The Arbitrarily Varying Wiretap Channel – Secret Randomness, Stability and Super-Activation

J. Nötzel[1,3], M. Wiese[2], H. Boche[1]

Electronic addresses: janis.noetzel@tum.de, moritzw@kth.se, boche@tum.de

[1] Lehrstuhl für Theoretische Informationstechnik, Technische Universität München, 80290 München, Germany.

[2] ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden.

[3]Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain.

April 4, 2016

Abstract

We define the common randomness assisted capacity of an arbitrarily varying channel (AVWC) when the Eavesdropper is kept ignorant about the common randomness. We prove a multi-letter capacity formula for this model. We prove that, if enough common randomness is used, the capacity formula can be given a single-shot form again.
We then consider the opposite extremal case, where no common randomness is available, and derive the capacity. It is known that the capacity of the system can be discontinuous under these circumstances. We prove here that it is still *stable* in the sense that it is continuous around its positivity points. We further prove that discontinuities can only arise if the legal link is symmetrizable and characterize the points where it is positive. These results shed new light on the design principles of communication systems with embedded security features.
At last we investigate the effect of super-activation of the message transmission capacity of AVWCs under the average error criterion. We give a complete characterization of those AVWCs that may be super-activated. The effect is thereby also related to the (conjectured) super-activation of the common randomness assisted capacity of AVWCs with an eavesdropper that gets to know the common randomness.
Super-activation is based on the idea of "wasting" a few bits of non-secret messages in order to enable provably secret transmission of a large bulk of data, a concept that may prove to be of further importance in the design of communication systems. In this work we provide further insight into this phenomenon by providing a class of codes that is capacity-achieving and does not convey any information to the Eavesdropper.

Contents

1 Introduction **2**

2 Notation and Definitions **9**
2.1 Notation and Conventions 9
2.2 Models and operational definitions 11

3 Main Results **17**

4 Proofs **24**
4.1 Technical definitions and facts 24
4.2 Proof of the converse part of Theorem 1 (coding theorem for C_{key}) 26
4.3 Proof of the direct part of Theorem 1 (coding theorem for C_{key}) 27
4.4 An intermediate result 30
4.5 Proof of Lemma 1 38
4.6 Proof of Theorems 2, 3 and 4 (properties of C_{S}) 40
4.7 Proof of Lemma 2 45
4.8 Proof of Theorem 5 (super-activation results) 46
4.9 Proof of Lemma 3 47

5 Appendix (auxiliary results and proofs) **47**

1 Introduction

Just like in our previous work [38], we investigate a model on the intersection between the two areas of secrecy and robust communication in information theory: the arbitrarily varying wiretap channel (AVWC). The communication scenario is depicted in Figure 1.
In this model, a sender (Alice) would like to send messages to a legitimate receiver (Bob) over a noisy channel. Involved into the scenario are two other parties: a jammer (James) who can actively influence the channel and a second but illegitimate receiver (Eve). Alice's and Bob's goal is to achieve reliable and secure communication:
First, Bob should be able to decode Alice's messages with high probability (with respect to the average error criterion) no matter what the input of James is.
Second, the mutual information between the messages and Eve's output should be close to zero. Again, this has to be the case no matter what the input of James is.
Like in our previous work, we add the option of Alice and Bob having access to perfect copies of the outcomes of a random experiment $\mathcal{G}$ (a source of common randomness). While in our previous work [38] we considered the case where Eve gets an exact copy of the outcomes received by Alice and Bob, we now extend our study to the case where Eve remains completely ignorant. The only party which has no access to $\mathcal{G}$ in all the scenarios we study is James. We call the capacities which we derive from the two scenarios the "correlated random coding mean secrecy capacity" if Eve has information about $\mathcal{G}$ and "secret common randomness assisted secrecy capacity" if Eve has no information about it. When no common randomness is present at all, we speak of the "uncorrelated coding secrecy capacity". For the sake of an extended discussion of secrecy criteria we also define a "capacity with public side-information" which is the data transmission benchmark for systems where Eve gets to know a part of the messages.
From now on, we use the label C_{S} for the uncorrelated coding secrecy capacity (when no shared randomness is available between Alice and Bob) and $C_{\text{S,ran}}^{\text{mean}}$ for the correlated random coding mean secrecy capacity (just as in our previous work [38] we restrict attention to the case where common randomness is used. To the reader which is not familiar with that work we apologize, as some of our results rely on that previous work). The secret common randomness assisted secrecy capacity is labelled C_{key} and the capacity with public side information C_{pp}. As is depcited in Figure 1, it is of vital importance that Eve cannot communicate to James.

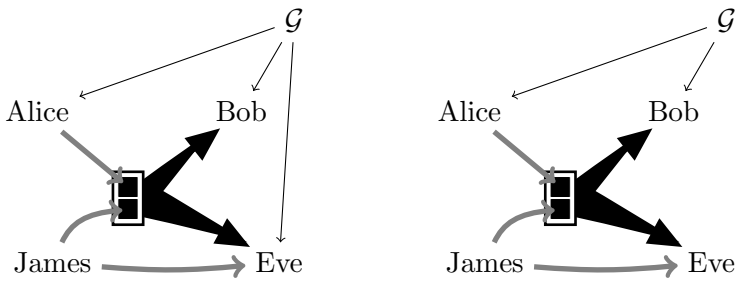

Figure 1: Secure coding schemes for correlated random coding (left) and secret common randomness assisted coding (right)

We give a unified treatment of the subject which allows us to observe the behaviour of the system while we change the amount of and the access to the common randomness: for common randomness set to zero one observes instabilities of the system (in the sense that the capacity is not a continuous function of the channel parameters anymore) and the effect of super-activation. Roughly speaking, two channels show super-activation when each of them cannot be used for a certain task (e.g. reliable communication under average error, maximal error or zero error criterion or, as in this work, secure communication) alone, but if a joint use is allowed the task becomes feasible. A more precise formulation is given in equations (5) to (7), while the definition is part of Definition 11 which is followed by a short discussion of super-activation in the scenario treated here. If common randomness is used between Alice and Bob but Eve gets to know it as well, it is known from the results in [38] that already small (a logarithmic number of bits, counted in block-length) amounts of common randomness resolve the instabilities (in the sense that the correlated random coding capacity is a continuous function of the channel parameters). It remains unknown whether super-activation is possible when common randomness is present, and this question is the content of Conjecture 1.
The full advantage from common randomness can only be gained if Eve is kept ignorant of it. If common randomness is used at a nonzero rate, this rate adds linearly to the capacity of the system. All the capacity formulas which can be proven to hold in the various nontrivial scenarios are given by multi-letter formulae. Only if the common randomness exceeds the maximal amount of information which can be leaked to Eve do we recover a single-letter description. At that point, the linear increase in capacity stops: In order to carve out these principal features of secure data transmission in a both exact and elegant mathematical framework we let the number n of channel uses go to infinity.
We will now sketch the connections of our work with some of the highlights and landmarks in the earlier literature. While we do not attempt to work in full rigour in the introduction, we will nonetheless gradually introduce some mathematical notation.
The probabilistic law which governs the transmission of codewords sent by Alice and jamming signals sent by James to Eve and Bob is, for n channel uses, given by

$$w^{\otimes n}(y^n|x^n,s^n)v^{\otimes n}(z^n|x^n,s^n)=\prod_{i=1}^{n}w(y_i|x_i,s_i)v(z_i|x_i,s_i). \tag{1}$$

Here, $s^n=(s_1,\ldots,s_n)$ are the inputs of James, $x^n=(x_1,\ldots,x_n)$ those of Alice and $z^n=(z_1,\ldots,z_n)$ the outputs of Eve, while $y^n=(y_1,\ldots,y_n)$ are received by Bob. All letters are

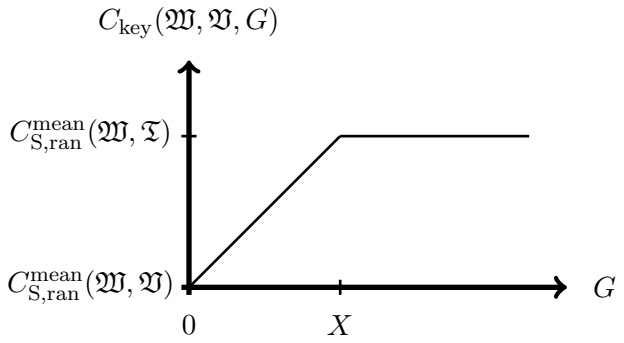

Figure 2: Scaling of secrecy capacity with the rate G of secret common randomness. It holds $X = C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W},\mathfrak{T}) - C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W},\mathfrak{V})$, where $\mathfrak{T}$ is defined below after equation (1).

assumed to be taken from finite alphabets. The action of the channel is, for each natural number n and therefore also as a whole, completely described by the pair (W,V) of matrices of conditional probabilities and this could rightfully be called an interference channel with non-cooperating senders and receivers. With respect to the historical development we will nonetheless prefer to use a description via the pair $(\mathfrak{W},\mathfrak{V}) = ((w(\cdot|\cdot,s))_{s\in\mathcal{S}},(v(\cdot|\cdot,s))_{s\in\mathcal{S}})$ and the label "AVWC".
This model has two important restrictions which are widely known: The case where $\mathfrak{V}$ does not convey any information about either one of its inputs is the arbitrarily varying channel (AVC). We will denote this special channel by $\mathfrak{T} = (T)$, where $t(z|x,s) = \frac{1}{|\mathcal{Z}|}$ for all z, x and s. Before we give some credit to the historical developments in the area, we would like to emphasize that the notion introduced in (1) extends to products of arbitrary channels from $\mathcal{X}_1$ to $\mathcal{Y}_1$ and $\mathcal{X}_2$ to $\mathcal{Y}_2$, let them be denoted W_1 and W_2 with respective transition probability matrices $(w_1(y|x))_{x\in\mathcal{X},y\in\mathcal{Y}}$ and $(w_2(y|x))_{y\in\mathcal{Y},x\in\mathcal{X}}$ as follows: The transition probability matrix of $W_1\otimes W_2$ is defined by $w(y_1,y_2|x_1,x_2) := w_1(y_1|x_1)\cdot w_2(y_2|x_2)$ (for all $x_1\in\mathcal{X}_1$, $x_2\in\mathcal{X}_2$, $y_1\in\mathcal{Y}_1$ and $y_2\in\mathcal{Y}_2$).
The notation then carries over to arbitrarily varying channels, where we set

$$\mathfrak{W}\otimes\mathfrak{W}' := (W_s\otimes W'_{s'})_{s\in\mathcal{S},s'\in\mathcal{S}'}. \tag{2}$$

The model of an arbitrarily varying channel has been introduced by Blackwell, Breiman and Thomasian [12] in 1960. They derived a formula for the capacity of an AVC with shared randomness-assisted codes under the average error criterion, and we will restrict our discussions to this criterion, although important nontrivial results concerning message transmission under the maximal error criterion have been obtained e.g. in [32, 1]. In [1] it was shown that an explicit formula for the (weak) capacity of an AVC under maximal error criterion would imply a formula for the zero-error capacity of a discrete memoryless channel. The latter problem is open now for half a century.
In [2], Ahlswede developed an elegant and streamlined method of proof that, together with the random coding results of [12], enabled him to prove the following: the capacity of an AVC (under the average error probability criterion) is either zero or equals its random coding capacity. This dichotomic behaviour is extended in the present work to the case where there is a (nontrivial) eavesdropper that has access to the shared randomness.
After the discoveries made in [2], an important open question was, when exactly the deterministic capacity with vanishing average error is equal to zero, and in some sense the corresponding

question for the AVWC is left open by us as well. In 1985, a first step towards a solution was made by Ericson [27], who came up with a sufficient condition that was proven to be necessary by Csiszar and Narayan [22] in 1989.
The condition which was developed by Ericson, called *symmetrizability*, reads as follows: An AVC $\mathfrak{W}$ is called symmetrizable if there is a set $(u(\cdot|x))_{x\in\mathcal{X}}$ of probability distributions on $\mathcal{S}$ such that for every $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$ we have

$$\sum_{s\in\mathcal{S}} u(s|x)w(y|x', s) = \sum_{s\in\mathcal{S}} u(s|x')w(y|x, s). \tag{3}$$

An arbitrarily varying channel $\mathfrak{W}$ that is symmetrizable cannot be used for reliable transmission of messages, as any input x can, at least in an average sense, be made to look as if it had been another input x'. An example for a symmetrizable AVC that cannot be used for reliable transmission of messages just by using one encoder-decoder pair but still has a positive capacity for correlated random codes was given in [12] and later used again in [2, Example 1]. This exemplary AVC also serves as an important ingredient to the super-activation results in [16] and is, as an important example, also to be found in Remark 7 of this document.
On the technical side, this work makes heavy use of the results that were obtained in the work [22] by extending one of their central results to the situation where Eve gets some information via V. Namely, we are able to prove the following: If $\mathfrak{W}$ is non-symmetrizable, then $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$ for all possible $\mathfrak{V}$. We do not attempt to give a necessary and sufficient condition for C_{S} to be positive, since a geometric characterization in the spirit of the symmetrizability condition 3 is not even known for the usual wiretap channel. Rather, when speaking about the wiretap channel one usually refers to the concept of "less noisy" channels that was developed in [21].
The wiretap channel has been studied widely in the literature. The analysis started with the celebrated work [40] of Wyner, an important follow-up work was [21], by Csiszar and Körner. While Wyner only treated the degraded case, Csiszar and Körner derived the capacity for the general discrete memoryless wiretap channel. The wiretap channel in the presence of common randomness which is kept secret from Eve (in this scenario, one could equally well speak of a secret key) was studied by Kang and Liu in [30].
In recent years there has been a growing interest in more elaborate models which combine insufficient channel state information with secrecy requirements. Probably the earliest publications which came to our attention are the work [34] by Liang, Kramer, Poor and Shamai and the paper [13] by Bloch and Laneman. Shortly after, the papers [9] and [10] by Bjelaković, Boche and Sommerfeld got published. The work [9] provides a lower bound on the secrecy capacity of the compound wiretap channel with channel state information at the transmitter that matches an upper bound on the secrecy capacity of general compound wiretap channels given provided in [34], establishing a full coding theorem in this case. Important contributions of the work [10] are a lower bound on what is called the "random code secrecy capacity" there, as well as a multi-letter expression for the secrecy capacity in the case of a best channel to the eavesdropper. The approach taken in this publication is closely related to the one taken in [10], but the use of different proof techniques enables us to provide much stronger results. An interesting parallel development is the work [29] by He, Khisti and Yener studies a two-transmitter Gaussian multiple access wiretap channel with multiple antennas at each of the nodes. A characterization of the secrecy degrees of freedom region under a strong secrecy constraint is derived.
A surprising result that was discovered only recently by Boche and Wyrembelski in [16] is

that of super-activation of AVWCs. We will explain this example in more detail in Remark 7. This effect was until then only known for information transmission capacities in quantum information theory, where it was proven by Smith and Yard in [36] that there exist channels which have the property that each of them alone has zero capacity but the two together have a positive capacity.
Before the work [16] this was assumed to be an effect which only shows up in quantum systems, where it was observed e.g. in [36].
The work [16] gave an explicit example of super-activation which we repeat in Remark 7, but a deeper understanding of the effect was not achieved. Based on our finer analysis, we are now able to provide the following results: First, we give a much clearer characterization of super-activation of the *uncorrelated*[1] coding secrecy capacity in Theorem 2. Second, and more for the sake of a clean discussion of coding and secrecy concepts, we define the capacity C_{pp} which explicitly keeps a part of the messages public (such that it may be that Eve is able to decode them). We do not attempt to give a further characterization of C_{pp} in this work, but we show that this capacity does as well show super-activation by use of the code concepts that were developed in [16]. Details are given in Subsection 2.2, together with the exact definition of C_{pp}.

We will now give a broad sketch of our results concerning C_{S} and $C_{\mathrm{S,ran}}^{\mathrm{mean}}$, before we start concentrating on C_{key}. It was proven in [38] that $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ is a continuous quantity, and while the statement may seem trivial at first sight, it becomes highly nontrivial when the following are taken into account:
There is at least no obvious way to deduce this statement directly just from the definition of the capacity, without first proving a coding result, and the latter route was taken in [38], where an explicit formula for $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ was found:

$$C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V}) = \lim_{n\to\infty} \frac{1}{n} \max_{p\in\mathcal{P}(\mathcal{U}_n)} \max_{U\in C(\mathcal{U}_n,\mathcal{X}^n)} \left(\min_{q\in\mathcal{P}(\mathcal{S})} I(p; W_q^{\otimes n}\circ U) - \max_{s^n\in\mathcal{S}^n} I(p; V_{s^n}\circ U) \right). \quad (4)$$

Explicit bounds on $|\mathcal{U}_n|$ were given as well. While one may argue that this is not an efficient description since one is forced to compute the limit of a series of convex optimization problems, it turns out to be an incredibly useful characterization in the following sense: First, it enables one to prove that $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ is a continuous function in the pair $(\mathfrak{W},\mathfrak{V})$ and this result was obtained in [38].
As has already been pointed out in [15], the continuous dependence of the performance of a communication system on the relevant system parameters is of central importance. To give just one example, consider recent efforts to build what is called "smart grids". Such systems do certainly have high requirements both concerning reliability and stability of the communication in order to avoid potentially damaging consequences for its users.
While it is very interesting from a mathematical point of view, it certainly comes as an unpleasant surprise then that C_{S} does not grant us the favour of being a continuous function of the

[1]Note that, due to the presence of an eavesdropper, it makes sense to allow the use of randomized encodings. Using, in such cases, the term "random code" is much too imprecise due to the potential presence of shared randomness between sender and receiver. Thus, we prefer to use the term uncorrelated codes. The random choice of codewords within an uncorrelated code represents lack of knowledge both for Eve and James. Analysing the case where James gains additional knowledge provides an interesting research opportunity, but care has to be taken when modelling the information flow from James to Eve.

channel. On the other hand, this casts a flashlight on the importance of distributed resources in communication networks - in this case the use of small amounts of common randomness. While one may now be tempted to think that the transmission of messages over AVWCs without the use of common randomness is a rather adventurous task, we are also able to prove that such a perception is wrong: Our analysis shows that C_{S} is continuous around its positivity points (this has been observed for classical-quantum arbitrarily varying channels in [15] already), and we are able to give an exact characterization of the discontinuity points as well. An example of a point of discontinuity has been given in [18].
Moreover, our characterization of discontinuity relies purely on the computation of functions which are *continuous* themselves, so that a calculation of such points is at least within reach also from a computational point of view.
Further, the deep interconnection between continuity and symmetrizability which shows up in our work enables us to give a characterization of pairs $(\mathfrak{W}_i, \mathfrak{V}_i)$ $(i = 1, 2)$ for which super-activation is possible only in terms of $C_{\mathrm{S,ran}}^{\mathrm{mean}}$. In order to be very explicit about super-activation, let us note the following:
The inequality

$$C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) \geqslant C_{\mathrm{S}}(\mathfrak{W}_1, \mathfrak{V}_1) + C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) \tag{5}$$

follows trivially from the definition of C. It is common to all notions of capacity which are known to the authors. In contrast, if the inequality

$$C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > C_{\mathrm{S}}(\mathfrak{W}_1, \mathfrak{V}_1) + C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) \tag{6}$$

holds, we speak of *super-additivity* and only if we can even find AVWCs $(\mathfrak{W}_1, \mathfrak{V}_1)$ and $(\mathfrak{W}_2, \mathfrak{V}_2)$ such that we have

$$C_{\mathrm{S}}(\mathfrak{W}_1, \mathfrak{V}_1) = C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) = 0, \qquad \text{but} \qquad C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0 \tag{7}$$

we speak of *super-activation*.
While it is clear from explicit examples in that super-activation of C_{S} is possible, it turns out in our work via Theorem 5 that the effect is connected to the super-activation of $C_{\mathrm{S,ran}}^{\mathrm{mean}}$, if the latter occurs. We would therefore like to take the opportunity of spurring future research by stating the following conjecture:

Conjecture 1. *There exist pairs* $(\mathfrak{W}_1, \mathfrak{V}_1)$ *and* $(\mathfrak{W}_2, \mathfrak{V}_2)$ *of (finite) AVWCs such that*

$$C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1, \mathfrak{V}_1) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1, \mathfrak{V}_1) = 0, \tag{8}$$

but

$$C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0. \tag{9}$$

An initial definition of objects such as $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ has been given in equation (2) and repeated again in Subsection 2.2. As a last introductory statement concerning super-additivity, let us mention the connection of super-activation to information transmission in networks: Consider two orthogonal channels in a mobile communication network. Not taking into account the issues on the physical layer, on may end up in a description of these channels via $\mathfrak{W}_1, \mathfrak{W}_2$ from Alice to Bob and $\mathfrak{V}_1, \mathfrak{V}_2$ from Alice to Eve. The surprising result then is that, while it may be

completely impossible to send information securely over each one of them, there exist coding schemes which enable Alice to send her information securely if both she and Bob have access to both $\mathfrak{W}_1$ and $\mathfrak{W}_2$!
We will argue later in Subsection 2.2 how this effect works for the capacity C_{pp}. While this capacity offers an insightful view on the topic, we nevertheless concentrate on the interplay between C_{S}, $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ and C_{key} in this work.

Let us now switch our attention to further results presented in this work. As mentioned already, we also extend earlier research to the case where *lots* of common randomness can be used (exponentially many random bits, to be precise) during our investigation of C_{key}. We do not dive into the issues arising when sub-exponentially many random bits are available, although the repeated appearance of the activating effect of common randomness in arbitrarily varying systems seems to deserve a closer study. Our method of proving the direct part does again yield nothing more than the statement that any number of random bits which scales asymptotically as $\mathrm{const.} + (1+\epsilon)\log(n)$ (for some $\epsilon > 0$) is sufficient for evading all issues which may arise from symmetrizable $\mathfrak{W}$.
Our restriction to positive rates G of common randomness allows us to give an elegant formula for C_{key} as follows: For every $G > 0$, it holds

$$C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G) = \min\{C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) + G, C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{T})\}. \tag{10}$$

Here, $\mathfrak{T}$ denotes the AVC consisting only of the memoryless "trash" channel T mapping every legal input x and jamming input s onto an arbitrary element of $\mathcal{Z}$ with equal probability ($t(z|s,x) = |\mathcal{Z}|^{-1}$). While the reader familiar with the topic would certainly have guessed the validity of a formula of this form it is worth noting that this formula is generally "hard to compute" in the sense that it requires one to calculate the limit in the formula (4) - as long as $G < C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{T}) - C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$. If this condition is not met, then $C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{T})$. Since the latter is the usual capacity of the AVC $\mathfrak{W}$, we conclude the following: If enough common randomness is available, the capacity of the system can be much more efficiently described - by a formula which does not require regularization anymore!
Again, a look into the area of quantum information theory shows a striking resemblance: The capacity formula for the usual memoryless quantum channel has been proven to be given by regularized quantities in the general cases, both for entanglement transmission and for message transmission. Without going into too much detail about quantum systems we cite here the work [23] by Devetak as our main reference underlining this statement, although this work has been both preceded and followed by important results dealing with the topic.
Apart from specific classes of quantum channels which were shown to have non-regularized capacity formulae [24] by Devetak and Shor, it has also been proven that the entanglement assisted capacity for message transmission over quantum channels is given by a one-shot formulae [8] by Bennet, Shor, Smolin and Thapliyal.
To the best of our knowledge, a quantification of the amount of entanglement assistance which is necessary in order to turn the capacity formula into a one-shot formula has not been given yet.

2 Notation and Definitions

This section contains notation, conventions, as well as operational definitions and technical definitions

2.1 Notation and Conventions

In the context presented in this work, every finite set will equivalently be called an alphabet. Such alphabets are denoted by script letters such as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. Every natural number $N \in \mathbb{N}$ defines a set $[N] := \{1, \ldots, N\}$. The set of all permutations on such $[N]$ is written S_N. The function $\exp : \mathbb{R} \to \mathbb{R}_+$ is defined with respect to base 2: $\exp(t) := 2^t$. The logarithm log is defined with respect to the same base. For any $c \in \mathbb{R}$ we define $|c|^+$ by setting $|c|^+ := c$ if $c > 0$ and $|c|^+ := 0$ otherwise. A function $f : \mathcal{A} \to \mathbb{R}$ is nonnegative ($f \geqslant 0$) if $f(a) \geqslant 0$ holds for all $a \in \mathcal{A}$. To each finite set $\mathcal{A}$ we associate the corresponding set $\mathcal{P}(\mathcal{A}) := \{p : \mathcal{A} \to [0,1] : p \geqslant 0, \sum_{a\in\mathcal{A}} p(a) = 1\}$ of probability distributions on $\mathcal{A}$. Each random variable A with values in $\mathcal{A}$ is associated to the unique $p \in \mathcal{P}(\mathcal{A})$ satisfying $\mathbb{P}(A = a) = p(a)$ for all $a \in \mathcal{A}$. An important subset of $\mathcal{P}(\mathcal{A})$ is the set of its extreme points. Every such extreme point is a point measure $\delta_a(a') := \delta(a, a')$ where $\delta(\cdot,\cdot)$ is the usual Kronecker-delta. The one-norm distance between two probability distributions $p, p' \in \mathcal{P}(\mathcal{A})$ is $\|p - p'\|_1 = \sum_{a\in\mathcal{A}} |p(a) - p'(a)|$.
The expectation of a function $f : \mathcal{A} \to \mathbb{R}$ with respect to a distribution $p \in \mathcal{P}(\mathcal{A})$ is written $\mathbb{E}_p f := \sum_{s\in\mathcal{A}} p(a)f(a)$ or, if p is clear from the context, simply $\mathbb{E}f$.
For each alphabet $\mathcal{A}$ and natural number $n \in \mathbb{N}$ we can build the corresponding product alphabet $\mathcal{A}^n := \mathcal{A} \times \ldots \times \mathcal{A}$, where $\times$ is the usual Cartesian product and there are exactly n copies of $\mathcal{A}$ involved in the definition of $\mathcal{A}^n$. The elements of $\mathcal{A}^n$ are denoted $a^n = (a_1, \ldots, a_n)$. Each such element gives rise to the corresponding empirical distribution or *type* $\bar{N}(\cdot|a^n) \in \mathcal{P}(\mathcal{A})$ defined via $N(a|a^n) := |\{i : a_i = a\}|$ and $\bar{N}(\cdot|a^n) := \frac{1}{n}N(\cdot|a^n)$. Given $\mathcal{A}$ and $n \in \mathbb{N}$, the set of all empirical distributions arising from an element $a^n \in \mathcal{A}^n$ is $\mathcal{P}_0^n(\mathcal{A}) := \{\bar{N}(\cdot|a^n) : a^n \in \mathcal{A}^n\}$. Each type $p \in \mathcal{P}_0^n(\mathcal{A})$ defines the *typical set* $T_p := \{a^n : \bar{N}(\cdot|a^n) = p(\cdot)\}$.
Channels are given by affine maps $W : \mathcal{P}(\mathcal{A}) \to \mathcal{P}(\mathcal{B})$. The set of channels is denoted $C(\mathcal{A}, \mathcal{B})$. Every channel is uniquely represented (and can therefore be identified with) its set $\{w(b|a)\}_{a\in\mathcal{A},b\in\mathcal{B}}$ of transition probabilities, which are defined via $w(b|a) := W(\delta_a)(b)$. It acts as

$$W(p) := \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} w(b|a)p(a)\delta_b, \tag{11}$$

where both $W(p) \in \mathcal{P}(\mathcal{B})$ and $\{\delta_b\}_{b\in\mathcal{B}} \subset \mathcal{P}(\mathcal{B})$ (another way of writing the above formula would be to set $W(p)(\cdot) = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} w(b|a)p(a)\delta_b(\cdot)$ or even $W(p)(y) = \sum_{a\in\mathcal{A}} w(b|a)p(a)$). As a shorthand, we may occasionally also write Wp to denote $W(p)$, in analogy to linear algebra (every channel is naturally associated to its representing stochastic matrix $(w(a|b))_{a,b}$ and can therefore be extended to a linear map on the appropriate vector spaces).
When operating on product alphabets such as $\mathcal{A} \times \mathcal{B}$ we define $p \otimes q \in \mathcal{P}(\mathcal{A} \times \mathcal{B})$ to be the distribution defined by $(p \otimes q)(a, b) := p(a)q(b)$. Correspondingly, $p^{\otimes n} \in \mathcal{P}(\mathcal{A}^n)$ is defined via $p^{\otimes n}(a^n) := \prod_{i=1}^n p(a_i)$. The same conventions hold for channels: if $V : \mathcal{P}(\mathcal{A}) \to \mathcal{P}(\mathcal{B})$ and $W : \mathcal{P}(\mathcal{A}') \to \mathcal{P}(\mathcal{B}')$, then $V \otimes W : \mathcal{P}(\mathcal{A} \times \mathcal{A}') \to \mathcal{P}(\mathcal{B} \times \mathcal{B}')$ is defined via its transition probabilities as $(v \otimes w)((b, b')|(a, a')) := v(b|a)w(b'|a')$ and the notation carries over to n-fold products $W^{\otimes n}$ of $W : \mathcal{P}(\mathcal{A}) \to \mathcal{P}(\mathcal{B})$ as before by setting $w^{\otimes n}(b^n|a^n) := \prod_{i=1}^n w(b_i|a_i)$.

For channels $W \in C(\mathcal{A}\times\mathcal{B},\mathcal{C})$ it will become important to derive a short notation for cases where one input remains fixed while the other is arbitrary. Such induced channels will be denoted, in case that this is unambiguously possible, by W_p where

$$W_p(\delta_a) := W(\delta_a \otimes p). \tag{12}$$

At times it will, in order to straighten out notation, also be necessary to write the transition probabilities as either $w_p(b|a)$ or even $w(b|a,p)$.
The Shannon entropy of $p \in \mathcal{P}(\mathcal{A})$ is $H(p) := -\sum_{a\in\mathcal{A}} p(A)\log p(a)$, the relative entropy between two probability distributions $p,q \in \mathcal{P}(\mathcal{A})$ is $D(p\|q) := \sum_{a\in\mathcal{A}} p(a)\log(p(a)/q(a))$, if $q(a) = 0 \Rightarrow p(a) = 0$ for all $a \in \mathcal{A}$, and $D(p\|q) := +\infty$, else.
Every $p \in \mathcal{P}(\mathcal{A})$ and channel $W : \mathcal{P}(\mathcal{A}) \to \mathcal{P}(\mathcal{B})$ define a joint random variable which we call (A,B) for the moment and which is defined via $\mathbb{P}((A,B) = (a,b)) = p(a)w(b|a)$ (for all $a \in \mathcal{A},\ b \in \mathcal{B}$). This enables us to use an equivalent formulation for the mutual information:

$$I(p;W) := I(A;B). \tag{13}$$

A more operational definition of this quantity can be achieved by noting that the distribution of (A,B) in this scenario arises from defining $p^{(2)} \in \mathcal{P}(\mathcal{A}\times\mathcal{A})$ by $p^{(2)}(a,a') := p(a)\cdot\delta_a(a')$ for all $a,a' \in \mathcal{A}$ - it then holds $\mathbb{P}((A,B) = (a,b)) = ((Id\otimes W)p^{(2)})(a,b)$ for all $a \in \mathcal{A}, b \in \mathcal{B}$. The operational interpretation of this probability distribution is that Alice observes the outcomes a of some random process. Given any such outcome, she makes one copy of it and sends that copy over to Bob via the channel W, keeping the original data with herself.
We will go one step further and define mutual information on pairs of sequences $a^n \in \mathcal{A}^n$, $b^n \in \mathcal{B}^n$, this time by defining a random variable (A,B) with values in $\mathcal{A}\times\mathcal{B}$ via $\mathbb{P}((A,B) = (a,b)) := \bar{N}(a,b|a^n,b^n)$ and then setting

$$I(a^n;b^n) := I(A;B). \tag{14}$$

In addition, we will need a suitable measure of distance between AVWCs. Our object of choice is the Hausdorff distance which we define as follows: For two channels $W,\tilde{W} \in C(\mathcal{A},\mathcal{B})$, set

$$\|W - \tilde{W}\| := \max_{a\in\mathcal{A}} \|W(\delta_a) - \tilde{W}(\delta_a)\|. \tag{15}$$

Now we define for a given $\mathfrak{W} = (W_s)_{s\in\mathcal{S}}$, and $\mathfrak{W}' = (W'_{s'})_{s'\in\mathcal{S}'}$

$$g(\mathfrak{W},\mathfrak{W}') := \max_{s\in\mathcal{S}} \min_{s'\in\mathcal{S}'} \|W_s - W'_{s'}\|.$$

Then we can ultimately define

$$d((\mathfrak{W},\mathfrak{V}),(\mathfrak{W}',\mathfrak{V}')) := \max\{g(\mathfrak{W}\otimes\mathfrak{V},\mathfrak{W}'\otimes\mathfrak{V}'), g(\mathfrak{W}'\otimes\mathfrak{V}',\mathfrak{W}\otimes\mathfrak{V})\}. \tag{16}$$

This is a metric on the set of finite-state AVWCs with the corresponding alphabets $\mathcal{A},\mathcal{B},\mathcal{C}$. Another ingredient in the following is the notion of the convex hull of a set of channels, which can for e.g. AVCs $\mathfrak{W} = (W_s)_{s\in\mathcal{S}}$ be defined as

$$\mathrm{conv}(\mathfrak{W}) := \left\{ W = \sum_{s\in\mathcal{S}} q(s)W_s : q \in \mathcal{P}(\mathcal{S}) \right\}. \tag{17}$$

At last, we would like to mention that for any given $W \in C(\mathcal{A},\mathcal{B})$, $a \in \mathcal{A}$ and subset $\mathcal{B}' \subset \mathcal{B}$ we use the notation

$$w(\mathcal{B}'|a) := \sum_{b\in\mathcal{B}'} w(b|a). \tag{18}$$

2.2 Models and operational definitions

At first, we give a formal definition of an arbitrarily varying channel. This extends our informal definition from the introduction, without any change in notation.

Definition 1 (AVWC)**.** *Let $\mathcal{X},\ \mathcal{Y},\ \mathcal{Z},\ \mathcal{S}$ be finite sets and for each $s \in \mathcal{S}$, let $W_s \in C(\mathcal{X}, \mathcal{Y})$ and $V_s \in C(\mathcal{X}, \mathcal{Z})$. Define $\mathfrak{W} := (W_s)_{s\in\mathcal{S}}$ and $\mathfrak{V} := (V_s)_{s\in\mathcal{S}}$. The corresponding arbitrarily varying wiretap channel is denoted $(\mathfrak{W}, \mathfrak{V})$. Its action is completely specified by the sequence $(\{W_{s^n}, V_{s^n}\}_{s^n\in\mathcal{S}^n})_{n\in\mathbb{N}}$, where $W_{s^n} := W_{s_1} \otimes \ldots \otimes W_{s_n}$ and $V_{s^n} := V_{s_1} \otimes \ldots \otimes V_{s_n}$.*

Remark 1. *The AVWC $(\mathfrak{W}, \mathfrak{V})$ can equivalently be represented by defining $W \in C(\mathcal{S} \times \mathcal{X}, \mathcal{Y})$ via $w(y|x, s) := w_s(y|x)$ and $V \in C(\mathcal{S} \times \mathcal{X}, \mathcal{Z})$ via $v(z|x, s) := v_s(y|x)$. We will use both representations interchangeably.*
Whenever necessary, we will (for $n \in \mathbb{N}$ and $q \in \mathcal{P}(\mathcal{S}^n)$) also use the abbreviations

$$W_q^{\otimes n} := \sum_{s^n\in\mathcal{S}^n} q(s^n)W_{s^n}, \qquad V_q^{\otimes n} := \sum_{s^n\in\mathcal{S}^n} q(s^n)V_{s^n}, \tag{19}$$

and the corresponding conditional probabilities are defined in the obvious way for all $x^n \in \mathcal{X}^n$, $y^n \in \mathcal{Y}^n$, $z^n \in \mathcal{Z}^n$ as

$$w_q^{\otimes n}(y^n|x^n) := W_q^{\otimes n}(\delta_{x^n})(y^n), \qquad v_q^{\otimes n}(z^n|x^n) := V_q^{\otimes n}(\delta_{x^n})(z^n). \tag{20}$$

Since a central part of our work is to study AVWCs under joint use, we have to carefully define what we mean here with "joint use". Let $(\mathfrak{W}_1, \mathfrak{V}_1)$ and $(\mathfrak{W}_2, \mathfrak{V}_2)$ be two AVWCs. Since state alphabets are finite in all of our work, we will without loss of generality assume that they have a joint state set $\mathcal{S}$. We then define

$$(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) := ((W_1(\cdot|\cdot, s) \otimes W_2(\cdot|\cdot, s'), V_1(\cdot|\cdot, s) \otimes V_2(\cdot|\cdot, s'))_{s,s'\in\mathcal{S}}. \tag{21}$$

We now come to a more "classic" topic: The definition of codes, rates and capacities. From the start, we will include the possibility of adding some extra variables like shared randomness or common randomness between Alice and Bob, but also the possibility for Alice to divide her message set into two parts: One which is to be kept secret from Eve and one which does not necessarily have to remain secret.
We introduce three different classes of codes, which are defined in the following and related to each other as follows: The class of shared randomness assisted codes contains those which use common randomness and these again contain the uncorrelated codes. Formal definitions are as follows:

Definition 2 (Shared randomness assisted code)**.** *A shared randomness assisted code $\mathcal{K}_n$ for the AVWC $(\mathfrak{W}, \mathfrak{V})$ consists of: a set $[K]$ of messages, two finite alphabets $[\Gamma], [\Gamma']$ and a set of stochastic encoders $e^\gamma \in C([K], \mathcal{X}^n)$ (one for every value $\gamma \in [\Gamma]$) together with a collection $((D_k^{\gamma'})_{k=1}^K)_{\gamma'=1}^{\Gamma'}$ of sets satisfying $\bigcup_{k=1}^K D_k^{\gamma'} \subset \mathcal{Y}^n$ and $D_k^{\gamma'} \cap D_{k'}^{\gamma'} = \varnothing$ for all $k \neq k'$ and for each γ'. In addition to that, there is a probability distribution $\mu \in \mathcal{P}([\Gamma] \times [\Gamma'])$. Every such code defines the joint random variables $\mathfrak{S}_{s^n} := (\mathfrak{K}_n, \mathfrak{K}'_n, \mathfrak{d}_n, \mathfrak{d}'_n, \mathfrak{Z}_{s^n}, \mathfrak{X}_n, \mathfrak{Y}_{s^n})$ ($s^n \in \mathcal{S}^n$) which are distributed according to*

$$\mathbb{P}(\mathfrak{S}_{s^n} = (k, k', \gamma, \gamma', z^n, x^n, y^n)) = \frac{1}{K}\mu(\gamma, \gamma')e^\gamma(x^n|k)\mathbb{1}_{D_{k'}^{\gamma'}}(y^n)w_{s^n}(y^n|x^n)v_{s^n}(z^n|x^n) \tag{22}$$

The average error of $\mathcal{K}_n$ is

$$\mathrm{err}(\mathcal{K}_n) = 1 - \max_{s^n \in \mathcal{S}^n} \sum_{k,\gamma,\gamma'=1}^{K,\Gamma,\Gamma'} \frac{\mu(\gamma,\gamma')}{K} \sum_{x^n \in \mathcal{X}^n} e^{\gamma}(x^n|k) w_{s^n}(D_k^{\gamma'}|x^n). \tag{23}$$

Definition 3 (Common randomness assisted code)**.** *A common randomness assisted code $\mathcal{K}_n$ for the AVWC $(\mathfrak{W}, \mathfrak{V})$ consists of: a set $[K]$ of messages, a set $[\Gamma]$ of values for the common randomness and a set of stochastic encoders $e^{\gamma} \in C([K], \mathcal{X}^n)$ (one for each element $\gamma \in [\Gamma]$), together with a collection $(D_k^{\gamma})_{k,\gamma=1}^{K,\Gamma}$ of subsets D_k^{γ} of $\mathcal{Y}^n$ satisfying $D_k^{\gamma} \cap D_{k'}^{\gamma} = \varnothing$ for all $\gamma \in [\Gamma]$, whenever $k \neq k'$. Every such code defines the joint random variables $\mathfrak{S}_{s^n} := (\mathfrak{K}_n, \mathfrak{K}'_n, \mathfrak{d}_n, \mathfrak{X}_n, \mathfrak{Y}_{s^n}, \mathfrak{Z}_{s^n})$ $(s^n \in \mathcal{S}^n)$ which are distributed according to*

$$\mathbb{P}(\mathfrak{S}_{s^n} = (k, k', \gamma, x^n, y^n, z^n)) = \frac{1}{\Gamma \cdot K} e^{\gamma}(x^n|k) \mathbb{1}_{D_{k'}^{\gamma}}(y^n) w_{s^n}(y^n|x^n) v_{s^n}(z^n|x^n) \tag{24}$$

The average error of $\mathcal{K}_n$ is

$$\mathrm{err}(\mathcal{K}_n) = 1 - \max_{s^n \in \mathcal{S}^n} \frac{1}{K \cdot \Gamma} \sum_{k,\gamma=1}^{K,\Gamma} \sum_{x^n \in \mathcal{X}^n} e^{\gamma}(x^n|k) w_{s^n}(D_k^{\gamma}|x^n). \tag{25}$$

For technical reasons we also define, for all state sequences s^n, the corresponding average success probability of the code by

$$d_{s^n}(\mathcal{K}_n) = \frac{1}{K \cdot \Gamma} \sum_{k,\gamma=1}^{K,\Gamma} \sum_{x^n \in \mathcal{X}^n} e^{\gamma}(x^n|k) w_{s^n}(D_k^{\gamma}|x^n). \tag{26}$$

One particularly interesting feature of AVCs is that it may be impossible to transmit any whatsoever small number of messages reliably from Alice to Bob without using shared randomness - but if one is willing to only spend a polynomial amount of common randomness, the capacity of the channel jumps to the maximally attainable value, an effect which was discovered in [2].

If a whole communication network is being utilized it may be possible to use one part of the network to establish common randomness between the legal parties (one could equally well speak of a secret key here) which is then used to send messages over another part of the system which may be symmetrizable. This idea was first established in [16]. In this work, we will give a more careful analysis of the underlying structure, an undertaking which motivates the following definition:

Definition 4 (Private/public code)**.** *A private/public code $\mathcal{K}_n$ for the AVWC $(\mathfrak{W}, \mathfrak{V})$ consists of: two sets $[K]$, $[L]$ of messages, an encoder $E \in C([K] \times [L], \mathcal{X}^n)$, and a collection $(D_{kl})_{k,l=1}^{K,L}$ of subsets of $\mathcal{Y}^n$ satisfying $D_{kl} \cap D_{k'l'} = \varnothing$ whenever $(k,l) \neq (k',l')$. Every such code defines the joint random variables $\mathfrak{S}_{s^n} := (\mathfrak{K}, \mathfrak{L}, \mathfrak{K}', \mathfrak{L}', \mathfrak{X}^n, \mathfrak{Y}_{s^n}, \mathfrak{Z}_{s^n})$ $(s^n \in \mathcal{S}^n)$ which are distributed according to*

$$\mathbb{P}(\mathfrak{S}_{s^n} = (k, l, k', l', x^n, y^n, z^n)) = \frac{1}{K \cdot L} e(x^n|k,l) \mathbb{1}_{D_{k'l'}}(y^n) w_{s^n}(y^n|x^n) v_{s^n}(z^n|x^n). \tag{27}$$

The average error of $\mathcal{K}_n$ is

$$\mathrm{err}(\mathcal{K}_n) = 1 - \max_{s^n \in \mathcal{S}^n} \sum_{k,l=1}^{K,L} \sum_{x^n \in \mathcal{X}^n} \frac{1}{K \cdot L} e(x^n|k,l) w_{s^n}(D_{k,l}|x^n). \tag{28}$$

With this definition we can formalize the idea of "wasting" a few bits in order to guarantee secret communication. We would like to compare this approach to the case of a compound channel, where a sender that knows the channel parameters can send pilot sequences to the receiver in order to let him estimate the channel. The pilot sequences do not carry information from sender to receiver. With such a scheme, a sender with state information can transmit at strictly higher rates than one without. The higher capacity is reached by "wasting" some transmissions for the estimation. Since the number of channel uses that have to be used for estimation grows only sub-exponentially in the number of channel uses, there is no negative impact on the message transmission rate in asymptotic scenarios.
In the case treated here it turns out that sending a small amount of non-secret messages is the key to increase the secrecy capacity in specific situations. We would like to extend the formal background of this idea by allowing for a joint transmission of secret and non-secret messages:

Definition 5 (Private/public coding scheme)**.** *A private/public coding scheme operating at rates* $(R_{\mathrm{pri}}, R_{\mathrm{pub}})$ *consists of a sequence* $(\mathcal{K}_n)_{n\in\mathbb{N}}$ *of private/public codes such that*

$$\lim_{n\to\infty} \mathrm{err}(\mathcal{K}_n) = 0, \tag{29}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log(K_n) = R_{\mathrm{pri}}, \tag{30}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log(L_n) = R_{\mathrm{pub}}, \tag{31}$$

$$\limsup_{n\to\infty} \max_{s^n\in\mathcal{S}^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}|\mathfrak{L}_n) = 0. \tag{32}$$

A more restricted class of codes arises when there is only one type of messages, which ought to be kept secret, and in addition allows the use of common randomness.

Definition 6 (Common randomness assisted coding scheme satisfying mean secrecy criterion)**.** *A common randomness assisted coding scheme satisfying the mean secrecy criterion operating at rate* R *consists of a sequence* $(\mathcal{K}_n)_{n\in\mathbb{N}}$ *of common randomness assisted codes such that*

$$\lim_{n\to\infty} \mathrm{err}(\mathcal{K}_n) = 0, \tag{33}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log(K_n) = R, \tag{34}$$

$$\limsup_{n\to\infty} \max_{s^n\in\mathcal{S}^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}|\mathfrak{b}_n) = 0. \tag{35}$$

Note that both Definition 5 and Definition 6 require the mutual information between the secret messages and the output at Eve's site to be small on average, either over the public messages or over the common randomness. One may argue that this is a somewhat weak criterion. In our earlier paper [38] we compared the capacity arising from the use of coding schemes under Definition 6 to a capacity derived under more severe requirements on the secrecy criterion. We were able to demonstrate that the respective capacities coincide. It is not known to us whether a more strict requirement in Definition 5 would lead to a different capacity.

Definition 7 (Secure uncorrelated coding scheme)**.** *A secure uncorrelated coding scheme operating at rate* R *consists of a sequence* $(\mathcal{K}_n)_{n\in\mathbb{N}}$ *of common randomness assisted codes with* $\Gamma_n = 1$

for all $n \in \mathbb{N}$ such that

$$\lim_{n\to\infty} \operatorname{err}(\mathcal{K}_n) = 0, \tag{36}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log(K_n) = R, \tag{37}$$

$$\limsup_{n\to\infty} \max_{s^n \in \mathcal{S}^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) = 0. \tag{38}$$

Definition 8 (Secure coding scheme with secret common randomness)**.** *A secure coding scheme with secret common randomness $\mathfrak{K}$ operating at rate R and using an amount $G_{\mathfrak{K}} > 0$ of common randomness consists of a sequence $\mathfrak{K} := (\mathcal{K}_n)_{n\in\mathbb{N}}$ of common randomness assisted codes with $\lim_{n\to\infty} \frac{1}{n} \log \Gamma_n = G_{\mathfrak{K}}$ such that*

$$\lim_{n\to\infty} \operatorname{err}(\mathcal{K}_n) = 0, \tag{39}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log(J_n) = R, \tag{40}$$

$$\limsup_{n\to\infty} \max_{s^n \in \mathcal{S}^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) = 0. \tag{41}$$

Remark 2. *The reader may wonder why the common randomness is only being quantified for secrecy schemes where the common randomness is kept secret. The reason for this becomes clear when reading [17], where it is proven that any shared randomness needed in order to achieve the correlated random coding mean secrecy capacity can always be assumed to not be larger than polynomially many bits of common randomness. These small amounts are not counted in the definition of the respective capacity. This result from [17] got applied in our earlier paper [38] as well.*

Since we completely restrict our analysis to the case where the system uses common randomness, we can spare a few indices to distinguish the different sources of external randomness:

Definition 9 (Secrecy capacities)**.** *Given $(\mathfrak{W}, \mathfrak{V})$, we define for every $G > 0$ the secret common randomness assisted secrecy capacity as*

$$C_{\text{key}}(\mathfrak{W}, \mathfrak{V}, G) := \sup \left\{ R : \begin{array}{l} \text{There exists secret common randomness assisted} \\ \text{coding scheme } \mathfrak{K} \text{ at rate } R \text{ with } G_{\mathfrak{K}} = G \end{array} \right\}. \tag{42}$$

The uncorrelated coding secrecy capacity and the correlated random coding mean secrecy capacity are defined just as in [38]:

$$C_{\text{S}}(\mathfrak{W}, \mathfrak{V}) := \sup \left\{ R : \begin{array}{l} \text{There exists a secure uncorrelated} \\ \text{coding scheme operating at rate } R \end{array} \right\} \tag{43}$$

$$C_{\text{S,ran}}^{\text{mean}}(\mathfrak{W}, \mathfrak{V}) := \sup \left\{ R : \begin{array}{l} \text{There exists a secure common randomness} \\ \text{assisted coding scheme satisfying the mean} \\ \text{secrecy criterion operating at rate } R \end{array} \right\}. \tag{44}$$

We refrain from defining the rate region for private and public messages in this work, and restrict ourselves to consider only the boundary of that region that arises from letting R_{pub} be arbitrarily small. This does for example allow us to transmit any finite number of messages, or numbers of messages that scale sub-exponentially in the number of channel uses.

Definition 10 (Private/public secrecy capacity)**.** *The private/public secrecy capacity is given by*

$$C_{\mathrm{pp}}(\mathfrak{W},\mathfrak{V}) := \sup\left\{R : \begin{array}{l}\text{There exists a private/public coding scheme at}\\ \text{rates } (R_{\mathrm{pub}},R_{\mathrm{pri}}) \text{ such that } R=R_{\mathrm{pri}}\end{array}\right\}. \tag{45}$$

The above definition explicitly allows for the super-activation strategy of [16] to be used, and shall be explained using this example first. Before we do so, let us give the formal definition of super-activation:

Definition 11 (Super-activation)**.** *Let* $(\mathfrak{W}_1,\mathfrak{V}_1)$ *and* $(\mathfrak{W}_2,\mathfrak{V}_2)$ *be AVWCs. Then* $(\mathfrak{W}_1,\mathfrak{V}_1)$*,* $(\mathfrak{W}_2,\mathfrak{V}_2)$ *are said to show super-activation if* $C_{\mathrm{S}}(\mathfrak{W}_1,\mathfrak{V}_1)=C_{\mathrm{S}}(\mathfrak{W}_2,\mathfrak{V}_2)=0$ *but* $C_{\mathrm{S}}(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)>0$*.*

Now set $\mathfrak{W}:=\mathfrak{W}_1\otimes\mathfrak{W}_2$ and $\mathfrak{V}:=\mathfrak{V}_1\otimes\mathfrak{V}_2$. In order to simplify the discussion, one may additionally set $\mathfrak{V}_2=\mathfrak{W}_2=(Id)$, where $Id\in C([2],[2])$ and assume that $\mathfrak{W}_1$ is symmetrizable but that $C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_1,\mathfrak{V}_1)=\alpha>0$. It follows that $C_{\mathrm{pp}}(\mathfrak{W}_1,\mathfrak{V}_1)=C_{\mathrm{pp}}(\mathfrak{W}_2,\mathfrak{V}_2)=0$, because of symmetrizability and since decoding of the messages that are sent via $(\mathfrak{W}_2,\mathfrak{V}_2)$ is possible without any error both for Bob *and* for Eve. These messages may therefore be treated as common randomness that is known by Eve. We know that already with the choice $L_n=n^2$ we have enough common randomness to remove any effect arising from symmetrizability of $\mathfrak{W}_1$. Since the code arising from the combination of sending and decoding public messages via (Id,Id) and private messages via $(\mathfrak{W}_1,\mathfrak{V}_1)$ is a coding scheme that fits under Definition 5, we get $C_{\mathrm{pp}}(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)\geqslant\alpha>0$.
That such a scheme does work as well when C_{S} is considered instead of C_{pp} can be understood as follows:
Let two AVWCs $(\mathfrak{W}_1,\mathfrak{V}_1)$ and $(\mathfrak{W}_2,\mathfrak{V}_2)$ be given. Let $\mathfrak{W}_1$ be symmetrizable, but such that $C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_1,\mathfrak{V}_1)=\alpha>0$. Since $\mathfrak{W}_1$ is symmetrizable we have $C_{\mathrm{S}}(\mathfrak{W}_1,\mathfrak{V}_1)=0$. If no additional resources are available, the surplus α in the common-randomness assisted secrecy capacity cannot be put to use. Let now $C_{\mathrm{S}}(\mathfrak{W}_2,\mathfrak{V}_2)=0$ but $C_{\mathrm{S}}(\mathfrak{W}_2,\mathfrak{T})=\beta>0$ ($\mathfrak{T}$ denotes the trash channel, so this just means that it is possible to reliably transmit messages over $\mathfrak{W}_2$). Then

$$C_{\mathrm{S}}(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)\geqslant\alpha>0 \tag{46}$$

and the reason for this effect is that (as before when we considered C_{pp}) a small amount of messages can be sent over $\mathfrak{W}_2$ and is then used as common randomness, therefore increasing the rate of messages that can be sent reliably over $\mathfrak{W}_1$ from zero to α. Of course, the messages sent over $\mathfrak{W}_2$ can be read by Eve. That this causes no problems with the security requirements can be seen by defining a toy-model where only two parallel channels with respective adversarially controlled channel states are considered. This is done as follows:
Let us define random variables $\mathfrak{R}_{s,\hat{s}}=(\mathfrak{M},\hat{\mathfrak{M}},\mathfrak{Z}_{1,s},\hat{\mathfrak{Z}}_{2,\hat{s}})$ where

$$\mathbb{P}(\mathfrak{R}=(m,\hat{m},z,\hat{z}))=\frac{1}{M}\frac{1}{\hat{M}}w_{1,s}(z|m,\hat{m})\hat{w}_{2,\hat{s}}(\hat{z}|\hat{m}) \tag{47}$$

and the channels $\{W_{1,s}\}_{s\in\mathcal{S}}$ and $\{\hat{W}_{\hat{s}}\}_{\hat{s}\in\hat{\mathcal{S}}}$ can be controlled by James separately. It is understood that m are the messages, whereas $\hat{m}$ are the values of the shared randomness that is distributed between Alice and Bob by using $\{\hat{W}_{2,\hat{s}}\}_{\hat{s}\in\hat{\mathcal{S}}}$. We assume that for some small $\epsilon\geqslant 0$ we have

$$\max_{\hat{s}\in\hat{\mathcal{S}}} I(\mathfrak{M};\hat{\mathfrak{Z}}_{2,\hat{s}}|\hat{\mathfrak{M}})\leqslant\epsilon. \tag{48}$$

Observe that $\hat{\mathfrak{Z}}_{2,\hat{s}}$ depends solely on $\hat{\mathfrak{M}}$ via the channel $\hat{W}_{2,\hat{s}}$ (this is where the fact that the two arbitrarily varying channels are used in parallel), so that the data processing inequality yields for every $s, \hat{s}$ that

$$I(\mathfrak{M}; \mathfrak{Z}_{1,s}, \hat{\mathfrak{Z}}_{2,\hat{s}}) \leqslant I(\mathfrak{M}; \mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}). \tag{49}$$

It is a consequence of the independence between $\mathfrak{M}$ and $\hat{\mathfrak{M}}$ that we can (for every s and $\hat{s}$) then continue this chain of estimates as follows:

$$\begin{aligned}
I(\mathfrak{M}; \mathfrak{Z}_{1,s}, \hat{\mathfrak{Z}}_{2,\hat{s}}) &\leqslant I(\mathfrak{M}; \mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) && (50)\\
&= H(\mathfrak{M}) + H(\mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) - H(\mathfrak{M}, \mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) && (51)\\
&= H(\mathfrak{M}, \hat{\mathfrak{M}}) + H(\mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) - H(\mathfrak{M}, \mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) - H(\hat{\mathfrak{M}}) && (52)\\
&= H(\mathfrak{M}|\hat{\mathfrak{M}}) + H(\mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) - H(\mathfrak{M}, \mathfrak{Z}_{1,s}, \hat{\mathfrak{M}}) && (53)\\
&= H(\mathfrak{M}|\hat{\mathfrak{M}}) + H(\mathfrak{Z}_{1,s}|\hat{\mathfrak{M}}) - H(\mathfrak{M}, \mathfrak{Z}_{1,s}|\hat{\mathfrak{M}}) && (54)\\
&= I(\mathfrak{M}; \mathfrak{Z}_{1,s}|\hat{\mathfrak{M}}) && (55)\\
&\leqslant \epsilon. && (56)
\end{aligned}$$

Thus it is clear that, in addition,

$$\max_{s\in\mathcal{S}, \hat{s}\in\hat{\mathcal{S}}} I(\mathfrak{M}; \mathfrak{Z}_{1,s}, \hat{\mathfrak{Z}}_{2,\hat{s}}) \leqslant \epsilon. \tag{57}$$

It is also evident that this argument ceases to hold true when the channels that are used for transmission of M and of $\hat{M}$ ar not orthogonal anymore. Our sketch indicates why the protocol developed in [16] is able to meet the secrecy requirement in Definition 7.

After we indicated why the super-activation protocol works we do now want to switch the topic and highlight a few connections to related problems and technical difficulties:

It is evident from the existing literature on AVCs [5], arbitrarily varying classical-quantum channels [14] and on the quantification of shared randomness [7, 28, 31, 41, 39] that the latter is not an easy task. A brief overview concerning the connections between quantification of shared randomness and arbitrarily varying channels has been given in [14]. Our focus here is on systems that use only common randomness in various different ways.

In our previous work [38] we developed a formula for $C_{\mathrm{S,ran}}^{\mathrm{mean}}$. The proof, extending the results established in [16] and [17], displays clearly that already amounts of common randomness which scale polynomially in the blocklength n are sufficient for achieving the full random capacity. Moreover, an exact quantification of the amount of shared randomness is not necessary when speaking about correlated random coding mean secrecy capacity. Either no shared randomness is allowed in the sense that $\Gamma_n = 1$ for all $n \in \mathbb{N}$ or else one allows arbitrarily large amounts of it but then only uses the above mentioned polynomial amount.

With the functions $G \mapsto C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G)$ the story is a different one, as the following interesting behaviour occurs: They are well-defined for all $G > 0$. However, when $G = 0$ they are not unambiguously defined anymore, as it is clearly possible to take e.g. a sequence $(\Gamma_n)_{n\in\mathbb{N}}$ such that $\Gamma_n = n^2$ for each $n \in \mathbb{N}$. In that case, $G = \lim_{n\to\infty} \frac{1}{n} \log \Gamma_n = 0$, but the amount of randomness is sufficient in the sense that for every $\epsilon > 0$ there exists a sequence $(\mathcal{K}_n)_{n\in\mathbb{N}}$ of codes which use only the common randomness Γ_n, operate at a rate $R_\epsilon = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) - \epsilon$ and

are both asymptotically secure and satisfy $\lim_{n\to\infty} \mathrm{err}(\mathcal{K}_n) = 0$. Thus, purely from the mathematical definition of $C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G)$, one would be tempted to set $C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, 0) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$. However, from the operational point of view this is unsatisfying: imagine taking the statement "no common randomness" literally, and therefore setting $\Gamma_n = 1$ for all $n \in \mathbb{N}$. Let $\mathfrak{W}$ be a symmetrizable AVC. In that case there is no chance to reliably transmit *any* whatsoever small amount of messages with $\Gamma_n = 1$ for all $n \in \mathbb{N}$ [27].
It is thus clear that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) = \lim_{G\to 0} C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G)$ holds, but that it at least seems to be a difficulty to give a both operationally meaningful and mathematically satisfying definition of $C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, 0)$ (see e.g. [37] for a possible approach to such type of problem).

A quantity which will be proved to be of importance during our proofs and when quantifying how close an AVC is to being symmetrizable is defined as follows: We let M_{fin} denote the set of all finite sets of elements of $C(\mathcal{X}, \mathcal{Y})$.

Definition 12. *The function $F : M_{\mathrm{fin}} \to \mathbb{R}_+$ is defined via setting, for each $\mathfrak{W}' = (W'(\cdot|\cdot, s))_{s\in\mathcal{S}} \in M_{\mathrm{fin}}$,*

$$F(\mathfrak{W}') := \max_{U\in C(\mathcal{X},\mathcal{S})} \min_{x\neq x'} \left\| \sum_{s\in\mathcal{S}} u(s|x) W'(\delta_{x'} \otimes \delta_s) - \sum_{s\in\mathcal{S}} u(s|x') W'(\delta_x \otimes \delta_s) \right\|_1 . \tag{58}$$

This function obviously has the property that for every AWVC $\mathfrak{W}'$, the statement $F(\mathfrak{W}') = 0$ is equivalent to $\mathfrak{W}'$ being symmetrizable.

3 Main Results

In this section we list our main results. We start with a coding theorem concerning the secret common randomness assisted secrecy capacity whose direct part is based on our Lemma 1 that we state directly afterwards. We continue with a second and even more delicate lemma, which is an extension of [22, Lemma 3] to AVWCs. This lemma (Lemma 2) is important: it provides a direct (coding) part for Theorem 2, which addresses the influence of the symmetrizability condition (3) on the capacity C_S and thereby relates it to $C_{\mathrm{S,ran}}^{\mathrm{mean}}$.
Our last result connects to the work [16], which showed a very surprising effect that has so far not been observed for classical information-carrying systems: super-activation. We give a precise characterization of the conditions which lead to super-activation in Theorem 5.

Theorem 1 (Coding Theorem for secret common randomness assisted secrecy capacity)**.** *Let $(\mathfrak{W}, \mathfrak{V})$ be an AVWC. For every $n \in \mathbb{N}$, set $\mathcal{U}_n := [|\mathcal{X}|^n]$. Define*

$$C^*(\mathfrak{W}, \mathfrak{V}) := \lim_{n\to\infty} \frac{1}{n} \max_{p\in\mathcal{P}(\mathcal{U}_n)} \max_{U\in C(\mathcal{U}_n,\mathcal{X}^n)} \left(\min_{q\in\mathcal{P}(\mathcal{S})} I(p; W_q^{\otimes n} \circ U) - \max_{s^n\in\mathcal{S}^n} I(p; V_{s^n} \circ U) \right). \tag{59}$$

It holds (with $\mathfrak{T} = (T)$ denoting the AVC consisting only of the memoryless channel that assigns the uniform output distribution to every input symbol),

$$C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G) = \min\{C^*(\mathfrak{W}, \mathfrak{V}) + G, C^*(\mathfrak{W}, \mathfrak{T})\} \tag{60}$$

Of course, $C^*(\mathfrak{W}, \mathfrak{T})$ is the capacity of the AVC $\mathfrak{W}$ under average error. This capacity has a single-letter description. Since the first argument in above minimum is not single letter, there is

room for speculation whether there is room for improvement in this characterization or, if not, for which value of G the description in terms of a single-letter quantity is possible and for which not. Apart from the complicated multi-letter form, an important take-away from the above formula is that the following is true:

Corollary 1. *For every $G > 0$, the function $(\mathfrak{W}, \mathfrak{V}) \mapsto C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G)$ is continuous.*

Remark 3. *If $G = 0$ in the sense that $\Gamma_n = 0$ for all $n \in \mathbb{N}$, then for all AVWCs $(\mathfrak{W}, \mathfrak{V})$ we know that $C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G)$ equals $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V})$.*

We are getting closer to the technical core of our work now. The next Lemma is essential to proving the direct part of Theorem 1. It quantifies how many messages L and how many different values Γ for the common randomness are needed in order to make the output distributions at Eve's site independent from the chosen message k.

Lemma 1. *For every $\tau > 0$ there exists a value $\nu(\tau) > 0$ and an $N_0(\tau)$ such that for all $n \geqslant N_0(\tau)$ and natural numbers K, L, Γ and type $p \in \mathcal{P}_0^n(\mathcal{X})$ there exist codewords $(\mathbf{x}_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ in $T_p \subset \mathcal{X}^n$ and decoding sets $D_{kl}^{\gamma} \subset \mathcal{Y}^n$ obeying $D_{kl}^{\gamma} \cap D_{k'l'}^{\gamma} = \varnothing$ if $(k,l) \neq (k',l')$, such that we have:*
If $\frac{1}{n}\log(K \cdot L) \leqslant \min_{q \in \mathcal{P}(\mathcal{S})} I(p; W_q) - \nu(\tau)$ and $\Gamma \geqslant 2^{n \cdot 5 \cdot \nu(\tau)}$ then

$$\min_{s^n} \sum_{\gamma=1}^{\Gamma} \frac{1}{\Gamma} \sum_{k,l=1}^{K,L} \frac{1}{K \cdot L} w_{s^n}(D_{kl}^{\gamma} | \mathbf{x}_{kl\gamma}) \geqslant 1 - 2^{-n \cdot \nu(\tau)}. \tag{61}$$

If $\frac{1}{n}\log(L \cdot \Gamma) \geqslant \max_q I(p; V_q) + \tau$ then

$$\max_{s^n, k} \left\| \frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L,\Gamma} v_{s^n}(\cdot | \mathbf{x}_{kl\gamma}) - \mathbb{E} v_{s^n}(\cdot | X^n) \right\|_1 \leqslant 2^{-n \cdot \nu(\tau)}, \tag{62}$$

where X^n is distributed according to $\mathbb{P}(X^n = x^n) := \frac{1}{|T_p|} \mathbb{1}_{T_p}(x^n)$ and the dependence of ν on τ is such that $\lim_{\tau \to 0} \nu(\tau) = 0$.

While Lemma 1 delivers the correct interplay between and scaling of the size of the numbers of secret messages K, the number of additional messages L that are just being sent in order to obfuscate Eve and the number of values for the (secret) common randomness Γ that are being used up in the process, it is insufficient for dealing with the case when Γ is set to one or is kept very small. For those cases where the secret or partially secret common randomness Γ is set to one for every number of channel uses, we have to deal with the symmetrizability properties of the legal link $\mathfrak{W}$ from Alice to Bob. Initial statements in that case are as follows:

Theorem 2 (Symmetrizability properties of C_{S})**.** *Let $(\mathfrak{W}, \mathfrak{V})$ be an AVWC.*

1. *If $\mathfrak{W}$ is symmetrizable, then $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) = 0$.*
2. *If $\mathfrak{W}$ is non-symmetrizable, then $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$.*

We now start to take on a slightly different point of view, under which the AVWC becomes an object that has some parameters which can be subject to changes. When considering practical deployment aspects, such a point of view is necessary as all information we may have gathered about the channel during for example a training phase may not be accurate enough to model the real-world behaviour. Thus one needs to understand whether a slight error in the parameters may lead to catastrophic events, and this is the content of our next theorem.

Theorem 3 (Stability of C_{S})**.** *Let $(\mathfrak{W},\mathfrak{V})$ be an AVWC. If $(\mathfrak{W},\mathfrak{V})$ satisfies $C_{\mathrm{S}}(\mathfrak{W},\mathfrak{V})>0$ then there is an $\epsilon>0$ such that for all $(\mathfrak{W}',\mathfrak{V}')$ satisfying $d((\mathfrak{W},\mathfrak{V}),(\mathfrak{W}',\mathfrak{V}'))\leqslant\epsilon$ we have $C_{\mathrm{S}}(\mathfrak{W}',\mathfrak{V}')>0$.*

However, despite the reassuring statement of Theorem 3, care has to be taken at some points, which are characterized below.

Theorem 4 (Discontinuity properties of C_{S})**.** *Let $(\mathfrak{W},\mathfrak{V})$ be an AVWC.*

1. *The function C_{S} is discontinuous at the point $(\mathfrak{W},\mathfrak{V})$ if and only if the following hold: First, $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V})>0$ and second $F(\mathfrak{W})=0$ but for all $\epsilon>0$ there is $\mathfrak{W}_\epsilon$ such that $d(\mathfrak{W},\mathfrak{W}_\epsilon)<\epsilon$ and $F(\mathfrak{W}_\epsilon)>0$.*

2. *If C_{S} is discontinuous in the point $(\mathfrak{W},\mathfrak{V})$ then it is discontinuous for all $\hat{\mathfrak{V}}$ for which $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W},\hat{\mathfrak{V}})>0$.*

Note that $F(\mathfrak{W})=0$ is equivalent to $\mathfrak{W}$ being symmetrizable - a property which is defined in the introduction in equation (3). The function F itself is the content of Definition 12.

The take-away from above Theorem is two-fold: First, it delivers a criterion for the finding of a point of discontinuity that only requires the validation that $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ (a continuous function) is nonzero in a specific point and the running of a convex optimization problem (calculation of F in that point). Second, it becomes clear that any discontinuity of the capacity C_{S} arises solely from effects that stem from the "legal" link $\mathfrak{W}$ - changing $\mathfrak{V}$ has no effect on discontinuity.

Corollary 2. *For every $\mathfrak{W}$, the function $\mathfrak{V}\mapsto C_{\mathrm{S}}(\mathfrak{W},\mathfrak{V})$ is continuous.*

Note that discontinuity is caused both by the legal link $\mathfrak{W}$ (see statement 1) and the link $\mathfrak{V}$ to Eve (statement 2), but depends on $\mathfrak{V}$ only insofar as the capacity $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W},\hat{\mathfrak{V}})$ has to stay above zero in order for a discontinuity to occur.
Theorem 4 also delivers an efficient way for calculating whether C_{S} is discontinuous in a specific point or not: One only needs to give a good-enough approximation of the continuous function $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ and then run a convex optimization in order to calculate $F(\mathfrak{W})$. Regarding future research, it may therefore be of interest to quantify the degree of continuity of the capacity of arbitrarily varying channels in those regions where it is continuous.

Remark 4. *It is necessary to request the existence of the $\mathfrak{W}_\epsilon$ in the first statement of Theorem 4, and an easy example why this is so is the following:*
Define $W_{i,\epsilon}\in C(\{1,2\},\{1,2,3\})$ for $i=1,2$ and $\epsilon\in[0,1/2]$ by

$$W_{1,\epsilon}:=\begin{pmatrix}0 & 1-\epsilon\\ \epsilon & 0\\ 1-\epsilon & \epsilon\end{pmatrix},\qquad W_{2,\epsilon}:=\begin{pmatrix}1-\epsilon & 0\\ \epsilon & 1-\epsilon\\ 0 & \epsilon\end{pmatrix}. \tag{63}$$

For every $\epsilon \in [0,1/2]$, these AVCs are symmetrizable with $u(1|1) = \epsilon/(1-\epsilon)$ and $u(1|2) = (1-2\cdot\epsilon)/(1-\epsilon)$. The reason for this is that for every $\epsilon \in [0,1/2]$ the convex sets $\mathrm{conv}(\{W_{1,\epsilon}(\delta_1), W_{2,\epsilon}(\delta_1)\})$ and $\mathrm{conv}(\{W_{1,\epsilon}(\delta_2), W_{2,\epsilon}(\delta_2)\})$ have non-empty intersections. It is also geometrically clear that for any $\epsilon \in (0,1/2)$, there will be a small vicinity of AVCs which share this property. Thus, around such a $\mathfrak{W}_\epsilon$, all other AVCs are symmetrizable as well and for every $\mathfrak{V}$ we therefore have both $C_{\mathrm{S}}(\mathfrak{W}_\epsilon, \mathfrak{V}) = 0$ and $C_{\mathrm{S}}(\mathfrak{W}', \mathfrak{V}) = 0$ whenever $d(\mathfrak{W}_\epsilon, \mathfrak{W}')$ is small enough.
It is additionally clear from [12] that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_0, \mathfrak{T}) > 0$ and that it is therefore (since $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ is continuous by the results of [38]) possible to choose $\mathfrak{V}$ and $\delta > 0$ such that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_0, \mathfrak{V}) > 0$, $C_{\mathrm{S}}(\mathfrak{W}_0, \mathfrak{V}) = 0$, and $C_{\mathrm{S}}(\mathfrak{W}', \mathfrak{V}) = 0$ whenever $d(\mathfrak{W}_\delta, \mathfrak{W}')$ is small enough.
It is easy to see that the AVC $\mathfrak{W}_0$ does not share this property: Although $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_0, \mathfrak{T}) > 0$ and $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_0, \mathfrak{T}) > 0$, it is easy to find explicit examples of AVCs $\mathfrak{W}'$ which are arbitrarily close to $\mathfrak{W}_0$ but are non-symmetrizable.

Of course, every whatsoever nice characterization of a set of interesting objects is pretty useless if the set turns out to be empty. Fortunately, it has been proven in [18] that the function mapping an AVC $\mathfrak{W}$ to its capacity has discontinuity points by explicit example.
Such an example is also given by $(\mathfrak{W}_0, \mathfrak{T})$ with $\mathfrak{W}_0$ taken from above.

Remark 5. *The capacity $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$ was quantified in [38]. It satisfies*

$$C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) = \lim_{G \to 0} C_{\mathrm{key}}(\mathfrak{W}, \mathfrak{V}, G) = C^*(\mathfrak{W}, \mathfrak{V}). \tag{64}$$

The proofs of Theorems 1 and Theorem 2 are carried out by providing coding strategies. The proof of the direct part of Theorem 2 extends the techniques from [22] by adding constraints on the random code that lead to it having additional security features. These features are quantified in the following Lemma:

Lemma 2. *For any $\tau > 0$ and $\beta > 0$, there exists a value $\nu(\tau) > 0$ and an $N_0(\tau)$ such that for all $n \geqslant N_0(\tau)$, natural numbers K, L, Γ satisfying $K \cdot L \geqslant 2^{n\cdot\tau}$ and type $p \in \mathcal{P}_0^n(\mathcal{X})$ satisfying $\min_{x:p(x)>0} p(x) \geqslant \beta$, there exist codewords $(\mathbf{x}_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ in $T_p \subset \mathcal{X}^n$, and a $c' > 0$ such that if $\Gamma^{-1} > \exp(-2^{n\cdot c'})$ and upon setting $R = \frac{1}{n}\log(K\cdot L)$ we have*

$$\max_{\gamma, x^n, s^n} |\{(k,l) : (x^n, \mathbf{x}_{kl\gamma}, s^n) \in T_{\bar{N}(\cdot|x^n, \mathbf{x}_{kl\gamma}, s^n)}\}| \leqslant 2^{n(|R - I(\mathbf{x}_{kl\gamma}; x^n, s^n)|^+ + \tau)} \tag{65}$$

$$\max_{\gamma, s^n} |\{(k,l) : I(\mathbf{x}_{kl\gamma}; s^n) \geqslant \tau\}| \leqslant K \cdot L \cdot 2^{-n\cdot\tau} \tag{66}$$

$$\max_{\gamma, s^n} \left| \left\{ (k,l,\gamma) : \begin{array}{l} \text{There is } (k',l',\gamma') \neq (k,l,\gamma) \text{ such that} \\ I(\mathbf{x}_{kl\gamma}; \mathbf{x}_{k'l'\gamma'}, s^n) - |R - I(\mathbf{x}_{kl\gamma}; s^n)|^+ > \tau \end{array} \right\} \right| \leqslant K \cdot L \cdot 2^{-n\cdot\tau/2} \tag{67}$$

$$\frac{\log L\cdot\Gamma}{n} \geqslant \max_{q \in \mathcal{P}(\mathcal{S})} I(p; V_q) + \tau \;\Rightarrow\; \max_{s^n, k} \left\| \frac{1}{L\cdot\Gamma} \sum_{l,\gamma=1}^{L,\Gamma} V_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \mathbb{E} V_{s^n}(\cdot|X^n) \right\|_1 \leqslant 2^{-n\cdot\nu(\tau)} \tag{68}$$

where X^n is distributed according to $\mathbb{P}(X^n = x^n) := \frac{1}{|T_p|}\mathbb{1}_{T_p}(x^n)$ and the dependence of ν on τ is such that $\lim_{\tau\to 0}\nu(\tau) = 0$.

Our intention was be to apply this Lemma to AVWCs for which the link between Alice and Bob is not symmetrizable. While Lemma 2 contains the possibility to use shared randomness

Γ, this is not necessary in the application intended by us in this work (we use it only with Γ set to one). The main reason for keeping Γ as a variable in our proof this is that it allows us to deliver a unified treatment of the whole topic, increases the generality of the Lemma and does not require much additional work.

Remark 6. *The properties* (65), (66) *and* (67) *of the code are identical to those stated in [22, Lemma 3]. This Lemma again is the main ingredient to the proof of [22] that non-symmetrizability (symmetrizability is defined in (3)) is sufficient for message transmission over AVCs if the average error criterion and non-randomized codes are used. Our strategy thus is to use the properties* (65), (66) *and* (67) *in Lemma 2 in order to ensure successful message transmission over the legal link, if $\mathfrak{W}$ is non-symmetrizable.*
The main tool used by Csiszar and Narayan for proving properties (65), (66) *and* (67) *of Lemma 2 in their work [22] was large deviation theory, and this is where we can make the connection to our work and prove the additional properties via application of the Chernoff-bound.*
Roughly speaking, this method of proof amounts to adding some additional requirements in a situation where any exponential number of additional requirements can be satisfied simultaneously.

When utilizing Lemma 2 (with $\Gamma = 1$) in the proof of Theorem 2 one sees that while reliable transmission is achieved via fulfillment of conditions (65), (66) and (67) in Lemma 2 if and only if the legal link $\mathfrak{W}$ is non-symmetrizable, the security of the communication can always be achieved by making L large enough. This implies that there are generic communication systems (AVWCs with a symmetrizable legal link) for which it is much easier to design codes that convey little information to Eve than codes which ensure robust communication.

In order to derive from Lemma 1 the connection between symmetrizability and the capacity C_S (which is the content of Theorem 2) it is necessary to prove not only achievability of quantities like e.g. $\min_q I(p; W_q) - \max_s I(p; V_s)$ but also of quantities like $\min_q I(p'; W_q^n \circ U) - \max_{s^n} I(p'; V_{s^n} \circ U)$ that involve multiple channel uses and pre-coding that is defined via the optimization problem (4). Such a process of adding pre-coding may unfortunately cause the AVWC arising from the concatenation of pre-coding and the original AVWC to be symmetrizable. This highly interesting interplay of pre-coding and symmetrizability is quantified in the next Lemma and the following example.

Lemma 3. *Let $\mathfrak{W}$ be an arbitrarily varying channel with input alphabet $\mathcal{A}$, output alphabet $\mathcal{B}$ and state set $\mathcal{S}$. Let $T \in C(\mathcal{A}', \mathcal{A})$ be a channel. Let $\mathfrak{W}'$ be the arbitrarily varying channel with input alphabet $\mathcal{A}'$, output alphabet $\mathcal{B}$ and state set $\mathcal{R}$ defined by $w'(b|a', s) := \sum_{a \in \mathcal{A}} w(b|a, s)t(a|a')$ (or, equivalently, via setting $W'_s := W_s \circ T$ for all $s \in \mathcal{S}$).*
If $\mathfrak{W}$ is symmetrizable then $\mathfrak{W}'$ is symmetrizable as well.

That, even for channels T whose associated matrix $(t(a|a')_{a' \in \mathcal{A}', a \in \mathcal{A}}$ has full range, the reverse implication "$\mathfrak{W}'$ is symmetrizable $\Rightarrow$ $\mathfrak{W}$ is symmetrizable" does not hold came as a surprise and is proven here by explicit example:

Example 1. *Define an AVC $\mathfrak{W} \subset C(\{x_1, x_2\}, \{1, 2, 3\})$ by setting*

$$w(\cdot|s_1, x_1) := \delta_1, \tag{69}$$
$$w(\cdot|s_2, x_1) := \delta_2, \tag{70}$$
$$w(\cdot|s_1, x_2) := 0.6\delta_1 + 0.2\delta_2 + 0.2\delta_3, \tag{71}$$
$$w(\cdot|s_2, x_2) := 0.1\delta_1 + 0.3\delta_2 + 0.6\delta_3, \tag{72}$$

where $\delta_i(j) = 1$ if and only if $i = j$ holds for $i, j \in [3]$. Then W is non-symmetrizable: The equation

$$\lambda w(\cdot|s_1, x_1) + (1-\lambda) w(\cdot|s_2, x_1) = \mu w(\cdot|s_1, x_2) + (1-\mu) w(\cdot|s_2, x_2) \tag{73}$$

cannot have a solution with $\lambda, \mu \in [0, 1]$ because δ_3 appears only on the right hand side and with strictly positive weights.

However, if we add pre-coding by a binary-symmetric channel N_p with parameter $p \in [0, 1]$ we obtain the new AVC $\mathfrak{W}'$ defined via $W'_s := W_s \circ N_p$ or, more concretely, by

$$w'(\cdot|s_1, x_1) = p\delta_1 + p'(0.2\delta_1 + 0.6\delta_2 + 0.2\delta_3) \tag{74}$$
$$w'(\cdot|s_2, x_1) = p\delta_2 + p'(0.1\delta_1 + 0.3\delta_2 + 0.6\delta_3) \tag{75}$$
$$w'(\cdot|s_1, x_2) = p'\delta_1 + p(0.6\delta_1 + 0.2\delta_2 + 0.2\delta_3) \tag{76}$$
$$w'(\cdot|s_2, x_2) = p'\delta_2 + p(0.1\delta_1 + 0.3\delta_2 + 0.6\delta_3) \tag{77}$$

where $p' := 1 - p$. We set $p = 0.4$. The equation

$$\lambda w'(\cdot|s_1, x_1) + (1-\lambda) w'(\cdot|s_2, x_1) = \mu w'(\cdot|s_1, x_2) + (1-\mu) w'(\cdot|s_2, x_2) \tag{78}$$

can be written out explicitly into three equations for the two parameters μ, λ. The solution is given by

$$\lambda = 31/37, \qquad \mu = 75/148. \tag{79}$$

This shows that $\mathfrak{W}'$ is symmetrizable. The situation is depicted as follows:

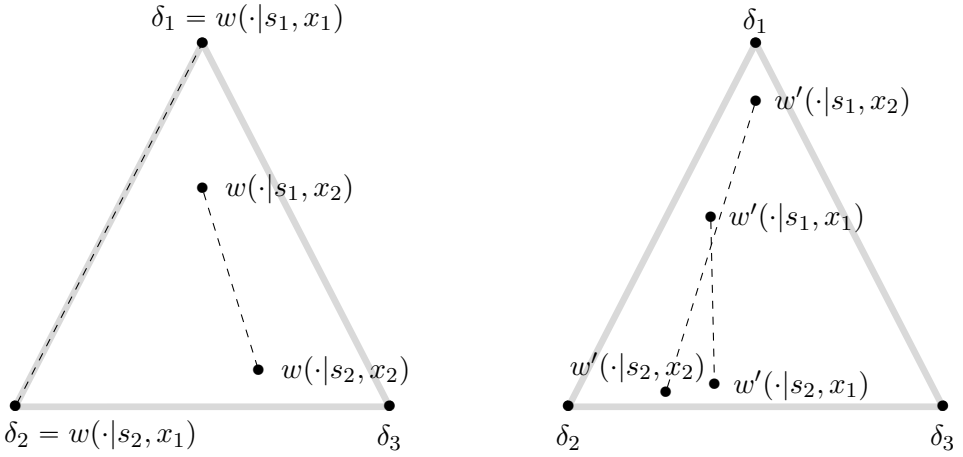

Figure 3: Light gray lines are the vertices of the probability simplex $\mathcal{P}(\{1, 2, 3\})$. The sets $\mathrm{conv}(\{w(\cdot|s_1, x_i), w(\cdot|s_2, x_i)\})$ where $i = 1, 2$ are displayed as dashed lines. The intersection of the dashed lines on the right shows that $\mathfrak{W}'$ is symmetrizable.

In order to derive the statement of Theorem 2 from Lemma 2 we can therefore not use a simple blocking strategy. Rather, we will present two methods of proof. The first employs a

reasoning along the lines of equations (48) until (57). This approach is based on the concept of using a few non-secret bits in order to guarantee secrecy for the actual data. While this is highly interesting from a practical point of view, it does not utilize the full strength of Lemma 2. This proof uses a set of public messages that can be read by Eve but not by James, secrecy is only obtained for the (exponentially larger) set of private messages.
Our second proof of Theorem 2 is based on lifting the optimal pre-codings for n channel uses to $n+1$ channel uses by using no pre-coding on the $(n+1)$th channel use. This type of pre-coding preserves non-symmetrizability. The second proof makes almost full use of the statements of Lemma 2, as we still set $\Gamma = 1$. No public messages are used in the construction of the code.

It remains an interesting open question whether, for n channel uses, the optimal channel U arising from the n-th term of the optimization problem (4) does in fact symmetrize $(W_{s^n})_{s^n \in \mathcal{S}^n}$ or not.

Our next result is potentially the most interesting in this work, since it sheds additional light on a rather new phenomenon: the super-activation of "the" secrecy capacity of AVWCs.

Theorem 5 (Characterization of super-activation of C_{S} via properties of $C_{\mathrm{S,ran}}^{\mathrm{mean}}$)**.** *Let $(\mathfrak{W}_i, \mathfrak{V}_i)_{i=1,2}$ be AVWCs.*

1. *If $C_{\mathrm{S}}(\mathfrak{W}_1, \mathfrak{V}_1) = C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) = 0$, then the estimate*

$$C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0 \tag{80}$$

   is true if and only if $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is not symmetrizable and $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0$. If $(\mathfrak{W}_i, \mathfrak{V}_i)_{i=1,2}$ can be super-activated it holds

$$C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2). \tag{81}$$

2. *There exist AVWCs which exhibit the above behaviour.*

3. *If $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ shows super-activation for $(\mathfrak{W}_1, \mathfrak{V}_1)$ and $(\mathfrak{W}_2, \mathfrak{V}_2)$, then C_{S} shows super-activation for $(\mathfrak{W}_1, \mathfrak{V}_1)$ and $(\mathfrak{W}_2, \mathfrak{V}_2)$ if and only if at least one of $\mathfrak{W}_1$ or $\mathfrak{W}_2$ is non-symmetrizable.*

4. *If $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ shows no super-activation for $(\mathfrak{W}_1, \mathfrak{V}_1)$ and $(\mathfrak{W}_2, \mathfrak{V}_2)$ then super-activation of C_{S} can only happen if $\mathfrak{W}_1$ is non-symmetrizable and $\mathfrak{W}_2$ is symmetrizable and $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1, \mathfrak{V}_1) = 0$ and $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_2, \mathfrak{V}_2) > 0$. The statement is independent of the specific labelling.*

Remark 7. *Of course for $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ to be non-symmetrizable, it has to be that at least one out of $\mathfrak{W}_1$, $\mathfrak{W}_2$ is non-symmetrizable.*
While Theorem 5 offers a complete characterization, it does not give any explicit examples - fortunately this has already been done in [16], where two AVWCs were used as follows: The first legal link is modeled by an AVC $\mathfrak{W}_1 = (W_{1,1}, W_{1,2})$ with input system for Alice being $\{1,2\}$ and output at Bob's site being $\{1,2,3\}$. The transition probabilities were given by

$$W_{1,1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{\top}, \qquad W_{1,2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}^{\top} \tag{82}$$

(note that assume that the columns of a matrix representing a channel sum up to one, not the rows!) and the first link to the eavesdropper by $\mathfrak{V}_1 = (V_1)$ (no influence from the jammer on

that link). For the purpose of this example, it would even be sufficient to let $\mathfrak{V}_1 = \mathfrak{T}$. This channel has the property that $\mathfrak{W}_1$ is symmetrizable. The second link was chosen to consist of two binary symmetric channels W_2, V_2 where W_2 was a degraded version of V_2, but both had nonzero capacity. Thus, $C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) = 0$ but nontheless it is possible to transmit (non-secret) messages via $\mathfrak{W}_2$. This example fits into the third class of pairs of AVWCs described in the above Theorem 5.
While this explicit example is very interesting, our analysis provides a more systematic analysis. Note that all our arguments only apply to the strong secrecy criterion. The weak secrecy criterion can be handled differently, and will be the scope of future work.

As a last point in this section, we would like to discuss connections between C_{pp} and C_{S}. At first, let us observe a similarity: The former shows super-activation if and only if the latter shows super-activation. To see this, we argue as follows: By definition, the class of codes which transmit public and private messages as defined in Definition 5 includes that according to Definition 7 where no public information is transmitted. Therefore it holds that $C_{\mathrm{pp}}(\mathfrak{W}, \mathfrak{V}) \geqslant C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V})$ for all AVWCs $(\mathfrak{W}, \mathfrak{V})$. Further, the definition of private/public codes according to Definition 5 is more narrow than the one of a common randomness assisted code according to Definition 3, so that every private/public code is at the same time also a common randomness assisted code. Especially, the public messages may be treated as if they were common randomness if $L = \Gamma$. Therefore, $C_{\mathrm{pp}}(\mathfrak{W}, \mathfrak{V}) > 0$ implies that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) > 0$ for all $(\mathfrak{W}, \mathfrak{V})$. We conclude from Theorem 2 that $C_{\mathrm{pp}}(\mathfrak{W}, \mathfrak{V}) > 0$ implies $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) > 0$ for all $(\mathfrak{W}, \mathfrak{V})$. This leads us to conclude that

$$\forall\ (\mathfrak{W}, \mathfrak{V}): \qquad C_{\mathrm{pp}}(\mathfrak{W}, \mathfrak{V}) > 0 \qquad \Leftrightarrow \qquad C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) > 0. \tag{83}$$

Let now C_{pp} show super-actication on $((\mathfrak{W}_1, \mathfrak{V}_1), (\mathfrak{W}_2, \mathfrak{V}_2))$. Then it follows from the statement in equation (83) that both $C_{\mathrm{S}}(\mathfrak{W}_1, \mathfrak{V}_1) = C_{\mathrm{S}}(\mathfrak{W}_2, \mathfrak{V}_2) = 0$ and $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) > 0$. Therefore, super-activation of C_{pp} implies super-activation of C_{S}.
In the reverse direction, let C_{S} show super-activation on the pair $((\mathfrak{W}_1, \mathfrak{V}_1), (\mathfrak{W}_2, \mathfrak{V}_2))$. From the statement in equation (3) we immediately see that C_{pp} shows super-activation as well.
Concerning differences, we note that a question we have to leave open is whether there could exist AVWCs $\mathfrak{W}, \mathfrak{V}$ such that $C_{\mathrm{pp}}(\mathfrak{W}, \mathfrak{V}) > C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V})$ holds.
This question is of huge practical importance, as it allows the quantification of the interplay between private and public communication in interfering networks when i.i.d. assumptions are not met, as is often the case.

4 Proofs

4.1 Technical definitions and facts

An important part of our results builds on the mathematical structure that was developed in [22]. The structure of the codes developed there builds on randomly sampling codewords which are all taken from one and the same set T_p. In our previous paper we used an approach that was built on sampling codewords according to some pruned distribution p' defined by $p'(x^n) := \frac{1}{p^{\otimes n}(T_{p,\delta}^n)} \mathbb{1}_{T_{p,\delta}^n} \cdot p^{\otimes n}(x^n)$ for some $p \in \mathcal{P}(\mathcal{X})$. The small deviation of p' from $p^{\otimes n}$ brings with it some benefits concerning asymptotic estimates. Since this work uses the outcomes of the earlier work [38], it would be desirable to use exactly the same technical approach.

However, due to the intended connection to [22], we cannot use p' in this work. Instead, we decided to use the same distribution as the one which was used in [22] which is in some sense further away from $p^{\otimes n}$. While this ensures seamless connectivity to [22], it also made us deviate (compared to for example our previous work [38]) from standard formulations in some other points, namely: We use a different notion of conditional typicality than before, and we define typical sets using the relative entropy rather than the one-norm.
This deviation is motivated by the fact that, for any finite alphabet $\mathcal{A}$ and $p \in \mathcal{A}$ as well as type $\bar{N} \in \mathcal{P}_0^n(\mathcal{A})$, we have $p^{\otimes n}(T_{\bar{N}}) = \mathrm{poly}(n)2^{-n \cdot D(p\|\bar{N})}$ for some polynomial poly in n. Therefore, defining typicality with respect to relative entropy gives the best control on the asymptotic behaviour of typical sets. All methods that use other distance measures for the definition of typicality need to relate these other measures to the relative entropy.
That the use of relative entropy is also elegant as compared to other methods can be seen as follows: Looking at [20, Definition 2.9] (which deals with typicality in the presence of channels and inputs to those channels) one sees an additional advantage of using relative entropy over using one-norm: defining typicality with respect to variational distance requires one to add additional assumptions which are not necessary when relative entropy is used, as the latter quantity can become infinite.
More precisely, let us assume we are given a channel $W \in C(\mathcal{A}, \mathcal{B})$ such and $(a^n, b^n) \in \mathcal{A}^n \times \mathcal{B}^n$ such that for one specific choice of a, b we have $N(a, b|a^n, b^n) > 0$ but $w(b|a) = 0$. Then b^n is not a typical output of the channel $w^{\otimes n}$ given that its input was a^n, since the probability that it is received when a^n as sent is zero:

$$\begin{aligned} 0 &\leqslant w^{\otimes n}(b^n|a^n) && (84)\\ &= \prod_{i=1}^{n} w(b_i|a_i) && (85)\\ &\leqslant \prod_{i:a_i=a,b_i=b} w(b_i|a_i) && (86)\\ &= w(b|a)^{N(a,b|a^n,b^n)} && (87)\\ &= 0^{N(a,b|a^n,b^n)} && (88)\\ &= 0. && (89) \end{aligned}$$

Excluding non-typical sequences is crucial for the derivation of lower bounds on cardinality of the conditionally typical set, for example. Thus, the above sequence b^n is excluded from the w-typical set given a^n explicitly in [20, Definition 2.9].
A notion using relative entropy captures this perfectly as well, but without necessitating the explicit exclusion: Let us assume that b^n is said to be w- typical given a^n iff $\Omega(a^n, b^n) := D(\bar{N}(\cdot|a^n, b^n)\|p_{AB})$ satisfies $\Omega(a^n, b^n) \leqslant \delta$ for some $\delta > 0$, where $p_{AB}(a, b) := \bar{N}(a)w(b|a)$. Then let a^n be given and b^n be such that there exists a, b such that $N(a, b|a^n, b^n) > 0$ but $w(b|a) = 0$. It follows $p_{AB}(a, b) = 0$ and therefore $D(\bar{N}(\cdot|a^n, b^n)\|p_{AB}) = \infty$, so that $\Theta(x^n, y^n) = \infty$ and hence b^n is not w-typical given a^n.
A brief look at robust typicality as defined in [35] shows that this quantity is also only related to relative entropy via inequalities.
Therefore, our definition achieves two goals: It connects in the most direct way to the relevant probability estimates and can be written down with minimal effort.
Thus, the sets which we will be using frequently in the following are, for arbitrary finite sets

$\mathcal{A}, \mathcal{B}, \mathcal{C}$, every $p \in \mathcal{P}(\mathcal{A})$, $\tilde{V} \in C(\mathcal{A} \times \mathcal{B}, \mathcal{C})$ and $\delta > 0$ defined as follows: for a given $(a^n, b^n) \in \mathcal{A}^n \times \mathcal{B}^n$ we define $p_{ABC} \in \mathcal{P}(\mathcal{A} \times \mathcal{B} \times \mathcal{C})$ via $p_{ABC}(a, b, c) := \bar{N}(a^n, b^n)\tilde{v}(c|a, b)$ and

$$T^n_{p,\delta} := \{a^n \in \mathcal{A}^n : D(\bar{N}(\cdot|a^n)\|p) \leqslant \delta\}, \tag{90}$$

$$T_{\tilde{V},\delta}(a^n, b^n) := \{c^n : D(\bar{N}(\cdot|a^n, b^n, c^n)\|p_{ABC}) \leqslant \delta\}. \tag{91}$$

These definitions are only valid for $\delta > 0$. Each $T_{V,\delta}(s^n, x^n)$ obeys the estimate

$$\tilde{v}^{\otimes n}(T_{\tilde{V},\delta}(a^n)|a^n) \geqslant 1 - 2^{-n\cdot\delta/2}, \tag{92}$$

for all $n \in \mathbb{N}$ such that $|\mathcal{A} \times \mathcal{B}|\frac{1}{n}\log(2n) \leqslant \delta$. We set, for every $p \in \mathcal{P}(\mathcal{X})$,

$$E(p) := \max_{q\in\mathcal{P}(\mathcal{S})} I(p; V_q) \qquad \text{and} \qquad B(p) := \min_{q\in\mathcal{P}(\mathcal{S})} I(p; W_q). \tag{93}$$

For the technical part of our proofs, the most important tool will be the Chernoff-Hoeffding bound:

Lemma 4. *Let b be a positive number. Let $Z_1, \ldots, Z_L$ be i.i.d. random variables with values in $[0, b]$ and expectation $\mathbb{E}Z_l = \nu$, and let $0 < \varepsilon < \frac{1}{2}$. Then*

$$\mathbb{P}\left\{\frac{1}{L}\sum_{l=1}^{L} Z_l \notin [(1 \pm \varepsilon)\nu]\right\} \leqslant 2\exp\left(-L \cdot \frac{\varepsilon^2 \cdot \nu}{3 \cdot b}\right), \tag{94}$$

where $[(1 \pm \varepsilon)\nu]$ denotes the interval $[(1 - \varepsilon)\nu, (1 + \varepsilon)\nu]$.

The proof can be found in [25, Theorem 1.1] and in [6].

4.2 Proof of the converse part of Theorem 1 (coding theorem for C_{key})

Main ingredients to this proof are Fano's inequality, data processing and almost-convexity of the entropy.

Proof of converse for secret common randomness assisted secrecy capacity. Let a sequence $\mathcal{K} = (\mathcal{K}_n)_{n=1}^{\infty}$ of common randomness-assisted codes be given such that for all $n \in \mathbb{N}$ we have

$$\min_{s^n\in\mathcal{S}^n} \frac{1}{\Gamma_n \cdot K_n} \sum_{\gamma,k=1}^{\Gamma_n,K_n} e^{\gamma}(x^n|k) w_{s^n}(D^{\gamma}_k|x^n) \geqslant 1 - \epsilon_n, \tag{95}$$

$$\max_{s^n\in\mathcal{S}^n} I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) \leqslant \epsilon_n, \tag{96}$$

and of course $\limsup_{n\to\infty} \epsilon_n = 0$. Set $R := \liminf_{n\to\infty} \frac{1}{n}\log K_n$, and $G := \lim_{n\to\infty} \frac{1}{n}\log \Gamma_n$. In addition to the random variable defined in Definition 3, consider $(\mathfrak{K}_n, \mathfrak{K}'_{q,n}, \mathfrak{d}_n)$ distributed as

$$\mathbb{P}((\mathfrak{K}_n, \mathfrak{Y}^n_q, \mathfrak{K}'_{q,n}, \mathfrak{d}_n) = (k, k', \gamma)) = \sum_{s^n\in\mathcal{S}^n} q^{\otimes n}(s^n)\mathbb{P}(\mathfrak{K}_n, \mathfrak{Y}_{q,n}, \mathfrak{K}'_n, \mathfrak{d}_n). \tag{97}$$

Then for all $n \in \mathbb{N}$, $q \in \mathcal{P}(\mathcal{S})$ and $s^n \in \mathcal{S}^n$ Fano's inequality implies

$$(1 - \epsilon_n)\log K_n \leqslant I(\mathfrak{K}_n; \mathfrak{K}'_{q,n}|\mathfrak{d}_n) - I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) + 1 + \epsilon_n. \tag{98}$$

We can apply the data processing inequality to get

$$(1-\epsilon_n)\log K_n \leqslant I(\mathfrak{K}_n;\mathfrak{Y}_q^n|\mathfrak{b}_n) - I(\mathfrak{K}_n;\mathfrak{Z}_{s^n}) + 1 + \epsilon_n, \tag{99}$$

and from e.g. Lemma 3.4 in [20] and independence of the random variables $\mathfrak{K}_n$ and $\mathfrak{G}_n$ it follows that the asymptotic scaling of the rate $\liminf_{n\to\infty}\frac{1}{n}\log K_n$ can be upper bounded through the following inequality:

$$(1-\epsilon_n)\log K_n \leqslant I(\mathfrak{K}_n;\mathfrak{Y}_q^n) - I(\mathfrak{K}_n;\mathfrak{Z}_{s^n}) + H(\mathfrak{b}_n) + 1 + \epsilon_n. \tag{100}$$

Since this estimate is valid for all $q\in\mathcal{P}(\mathcal{S})$ and $s^n\in\mathcal{S}^n$ we get

$$\log K_n \leqslant \frac{1}{1-\epsilon_n}\left(\min_{q\in\mathcal{P}(\mathcal{S})} I(\mathfrak{K}_n;\mathfrak{Y}_q^n) - \max_{s^n\in\mathcal{S}^n} I(\mathfrak{K}_n;\mathfrak{Z}_{s^n})\right) + \frac{1+\epsilon_n}{1-\epsilon_n} + \frac{\log\Gamma_n}{1-\epsilon_n}. \tag{101}$$

Define the distribution $p\in\mathcal{P}([K_n])$ and the channel $U\in C([K_n],\mathcal{X}^n)$ by

$$p(k) = \frac{1}{K_n}, \qquad U(x^n|k) := \sum_{\gamma=1}^{\Gamma_n}\frac{1}{\Gamma}e^{\gamma}(x^n|k) \qquad (k\in[K_n],\ x^n\in\mathcal{X}^n). \tag{102}$$

Then we arrive at

$$\log K_n \leqslant \frac{1}{1-\epsilon_n}\left(\min_{q\in\mathcal{P}(\mathcal{S})} I(p;W_q^{\otimes n}\circ U) - \max_{s^n\in\mathcal{S}^n} I(p;V_{s^n}\circ U)\right) + \frac{1+\epsilon_n}{1-\epsilon_n} + \frac{\log\Gamma_n}{1-\epsilon_n}. \tag{103}$$

Of course, we can obtain a more relaxed upper bound by optimizing over all $p\in\mathcal{P}([K_n])$ and $U\in C([K_n],\mathcal{X}^n)$. We then obtain (since $K_n\leqslant|\mathcal{X}^n|$ for every reliably working code and, therefore, $\mathcal{P}([K_n])\subset\mathcal{P}([|\mathcal{X}^n|])$ under the standard embedding $[K_n]\subset[|\mathcal{X}|^n]$) by further increasing the size of the input alphabet from K_n to $|\mathcal{X}|^n$ with $\mathcal{U}_n := [|\mathcal{X}|^n]$ that

$$R \leqslant \lim_{n\to\infty}\frac{1}{n}\max_{p\in\mathcal{U}_n}\max_{U\in C(\mathcal{U}_n,\mathcal{X}^n)}\left(\min_{q\in\mathcal{P}(\mathcal{S}^n)} I(p;W_q^{\otimes n}\circ U) - \max_{s^n\in\mathcal{S}^n} I(p;V_{s^n}\circ U)\right) + G. \tag{104}$$

As it has been proven in [38] that the capacity $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ equals the leftmost part in the above sum we have proven the desired result.
Another obvious bound on the capacity arises by ignoring all security issues: since $\mathcal{K}$ ensures an asymptotically perfect transmission, we have

$$\lim_{n\to\infty}\frac{1}{n}\log K_n \leqslant \max_{p\in\mathcal{P}(\mathcal{X})}\min_{q\in\mathcal{P}(\mathcal{S})} I(p;W_q). \tag{105}$$

This establishes the converse part of the coding theorem. □

4.3 Proof of the direct part of Theorem 1 (coding theorem for C_{key})

Let $G>0$ be given. Define $p := \arg\max_{p\in\mathcal{P}(\mathcal{X})}(B(p)-E(p))$. Set $G' := \max\{E(p),G\}$. Intuitively speaking, this is the amount of common randomness which can be put to use in the obfuscation of Eve. Choose a $\tau>0$ such that $\nu(\tau)$ from Lemma 1 satisfies $\nu(\tau)<G'$. Let $n\in\mathbb{N}$ be so that for all $n\geqslant N$ there is $p_n\in\mathcal{P}_0^n(\mathcal{X})$ such that $|B(p_n)-B(p)|\leqslant\max\{\tau,\nu(\tau)\}$ and $|E(p_n)-E(p)|\leqslant\max\{\tau,\nu(\tau)\}$. This can be achieved by approximating p through types p_n

via Lemma 8 and since both B and E are continuous functions. Take three sequences $(K_n)_{n=1}^{\infty}$, $(L_n)_{n=1}^{\infty}$, $(\Gamma_n)_{n=1}^{\infty}$ of natural numbers. Without loss of generality, we can ensure that $(\Gamma_n)_{n\in\mathbb{N}}$ satisfies both $\Gamma_n \leqslant 2^{n\cdot G'}$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \frac{1}{n}\log\Gamma_n = G'$. Let now $n \in \mathbb{N}$ satisfying $n \geqslant N$ be fixed but large enough such that in addition

$$E(p) - G' + 4\tau \geqslant \frac{1}{n}\log(L_n) \geqslant E(p) - G' + 2\tau, \tag{106}$$

$$B(p) - E(p) + G' - 4(\tau + \nu(\tau)) \geqslant \frac{1}{n}\log(K_n) \geqslant B(p) - E(p) + G' - 2(\tau + \nu(\tau)) \tag{107}$$

be satisfied, for all large enough $n \in \mathbb{N}$. This implies both

$$\begin{aligned}\frac{1}{n}\log(K_n \cdot L_n) &\leqslant B(p) - E(p) + G' - 4(\tau + \nu(\tau)) + E(p) - G' + 4\tau &(108)\\ &= B(p) - 4\nu(\tau) &(109)\\ &\leqslant B(p_n) - \nu(\tau) &(110)\end{aligned}$$

and

$$\frac{1}{n}\log(L_n \cdot \Gamma_n) \geqslant E(p) + 2\tau \geqslant E(p_n) + \tau. \tag{111}$$

Asymptotically, we also have this yields

$$\liminf_{n\to\infty} \frac{1}{n}\log(K_n) \geqslant B(p) - E(p) + G - 4\cdot(\tau + \nu(\tau)). \tag{112}$$

At the same time, the prerequisites of Lemma 1 are met such that a reliable sequence of codes exists which is also secure with respect to $\|\cdot\|_1$: For all large enough $n \in \mathbb{N}$ we have

$$\min_{s^n} \sum_{\gamma=1}^{\Gamma} \frac{1}{\Gamma} \sum_{k,l=1}^{K,L} \frac{1}{K\cdot L} w_{s^n}(D_{kl}^{\gamma}|\mathbf{x}_{kl\gamma}) \geqslant 1 - 2^{-n\cdot\nu(\tau)}, \tag{113}$$

$$\max_{s^n,k} \left\| \frac{1}{L\cdot\Gamma} \sum_{l,\gamma=1}^{L,\Gamma} v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \mathbb{E}v_{s^n}(\cdot|X^n) \right\|_1 \leqslant 2^{-n\cdot\nu(\tau)}. \tag{114}$$

It can already be seen that this yields reliable communication at any rate which is strictly below $B(p) - E(p) + G$ - we proved the achievability of rates close enough to $B(p) - E(p) + G$, but it is clear that time sharing between a trivial strategy where only one codeword is being transmitted (which is then automatically perfectly secure) and the strategy which was proven to work in the above will show achievability of all other rates $R \in [0, |B(p) - E(p) + G|^+]$. That we also get secure communication can be seen as follows: From [20, Lemma 2.7] we know that our exponential bound (62) asymptotically leads to fulfillment of the strong secrecy criterion.
We have thus proven that, for each $\tau' > 0$, the number

$$\max_{p\in\mathcal{P}(\mathcal{X})} \left(\min_{q\in\mathcal{P}(\mathcal{S})} I(p; W_q) - \max_{q\in\mathcal{P}(\mathcal{S})} I(p; V_q) \right) + G - \tau' \tag{115}$$

is an achievable rate. We now proceed by adding channels U at the sender and using blocks of the original channels together: Since we now know that, for every $r \in \mathbb{N}$, $G > 0$ and $\delta > 0$,

$p \in \mathcal{P}(\mathcal{U}_r)$ where $U_r := [|\mathcal{X}|^r]$ and $U \in C(\mathcal{U}_r, \mathcal{X}^r)$ there exist sequences $\mathcal{K} = (\mathcal{K}_m)_{m=1}^{\infty}$ such that for every $s^{r\cdot m} \in (\mathcal{S}^r)^m = \mathcal{S}^{r\cdot m}$ we have

$$\frac{1}{K_m}\frac{1}{\Gamma_m}\sum_{k=1}^{K_m}\sum_{\gamma=1}^{\Gamma_m}\sum_{x^{r\cdot m}} e(x^{r\cdot m}|k,\gamma) w_{s^{m\cdot r}}(D_k^{\gamma}|x^{r\cdot m}) \geqslant 1-\epsilon_m, \tag{116}$$

where $\mathbf{x}_{k,\gamma} \in U_r^m$ are codewords (each $x_{k,\gamma,i}$ is an element of $\mathcal{U}_r$) for $(W_{s^r} \circ U)_{s^r \in \mathcal{S}^r}$, and the stochastic encoder is $e(x^{r\cdot m}|k,\gamma) = \prod_{i=1}^{m} u(x_{ij}|x_{k,\gamma,i})$ for $x^{r\cdot m}$ and it holds that

$$\liminf_{m\to\infty} \frac{1}{m}\log K_m \geqslant \min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p; W_q \circ U) - \max_{s^r\in\mathcal{S}^r} I(p; V_{s^r}\circ U) + r\cdot G - \delta. \tag{117}$$

We can define values $t_n \in \{0,\dots,r-1\}$ by requiring $n = m\cdot r + t_n$ for them to hold for some suitably chosen $m = m(n) \in \mathbb{N}$. This quantity satisfies $-1 + n/r \leqslant m(n) \leqslant n/r$. For every $n \in \mathbb{N}$ we then define new decoding sets by

$$\hat{D}_k^{\gamma} := D_k^{\gamma} \times \mathcal{Y}^{t_n} \tag{118}$$

and new codewords by setting for some arbitrary but fixed x^{t_n}

$$\hat{x}_{k\gamma} := (x_{k\gamma}, x^{t_n}). \tag{119}$$

From the choice of codewords and the decoding rule it is clear that this code is asymptotically reliable. The asymptotic number of codewords (mind that $\hat{K}_n = K_{m(n)}$) calculated and normalized with respect to n, is

$$\begin{aligned}
\liminf_{n\to\infty}\frac{1}{n}\log\hat{K}_n &= \liminf_{n\to\infty}\frac{1}{m(n)\cdot r + t_n}\log K_{m(n)} && (120)\\
&\geqslant \liminf_{n\to\infty}\frac{1}{r}\cdot\frac{1}{m(n)+1}\log K_{m(n)} && (121)\\
&= \liminf_{n\to\infty}\frac{1}{r}\cdot\frac{1}{m(n)}\cdot\frac{m(n)}{m(n)+1}\cdot\log K_{m(n)} && (122)\\
&= \frac{1}{r}\liminf_{n\to\infty}\frac{1}{m(n)}\cdot\log K_{m(n)} && (123)\\
&= \frac{1}{r}\left(\min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U) + r\cdot G - \delta\right). && (124)
\end{aligned}$$

To see that every number $C^*(\mathfrak{W},\mathfrak{V}) - \epsilon$ is an achievable rate, take r, U and p such that

$$C^*(\mathfrak{W},\mathfrak{V}) - \epsilon/2 \leqslant \frac{1}{r}\left(\min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U)\right). \tag{125}$$

This is possible since in [38] it was (in addition to the equality $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\cdot,\cdot) = C^*(\cdot,\cdot)$) proven that

$$C^*(\mathfrak{W},\mathfrak{V}) = \lim_{r\to\infty}\frac{1}{r}\max_{p\in\mathcal{P}(\mathcal{U}_n)}\max_{U_n\in C(\mathcal{U},\mathcal{X}^n)}\left(\min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U)\right). \tag{126}$$

We set $\delta = r \cdot \epsilon/4$. Then from our preceding arguments it becomes clear that there is a sequence $\hat{\mathcal{K}}$ of asymptotically reliable codes at an asymptotic rate

$$\liminf_{n\to\infty} \frac{1}{n} \log \hat{K}_n \geqslant \frac{1}{r} \left(\min_{q \in \mathcal{P}(\mathcal{S}^r)} I(p; W_q^m \circ U) - \max_{s^r \in \mathcal{S}^r} I(p; V_{s^r} \circ U) + r \cdot G - r \cdot \epsilon/4 \right) \tag{127}$$

$$\geqslant \frac{1}{r} \left(\min_{q \in \mathcal{P}(\mathcal{S}^r)} I(p; W_q^m \circ U) - \max_{s^r \in \mathcal{S}^r} I(p; V_{s^r} \circ U) + r \cdot G \right) - \epsilon/4 \tag{128}$$

$$\geqslant C^*(\mathfrak{W}, \mathfrak{V}) + G - \epsilon/2 - \epsilon/4 \tag{129}$$

$$\geqslant C^*(\mathfrak{W}, \mathfrak{V}) + G - \epsilon. \tag{130}$$

This proves the direct part of the coding theorem.

4.4 An intermediate result

We now have to prove the core results from which all the other statements can be deduced. The idea of proof will be to make a random selection of the codewords $\mathbf{x}_{kl\gamma}$ where k are the messages, l are non-secret messages which are only being sent in order to obfuscate the received signal at Eve, and γ are the values of the common randomness. When applying the results to AVWCs, the decoder is the one defined in [22] whenever we study C_{S} and is defined here according to our needs for the study of C_{key}.
We define events $E_1, \ldots, E_5$ which describe certain desirable properties of our codewords, in dependence of $(\mathfrak{W}, \mathfrak{V})$ and the numbers K, L, Γ of available indices k, l, γ. We then use Chernoff bounds. This guarantees that the random selection of codewords has each single property we would like them to have with probability lower bounded by $1 - \exp(-2^{nc})$ for some positive constant $c > 0$ and all large enough n under some conditions on Γ, L and K which of course depend on $(\mathfrak{W}, \mathfrak{V})$ as well. Application of a union bound then reveals the existence of one particular choice of codewords that has all the desired properties simultaneously.
Using exactly this method of proof, Csiszar and Narayan [22, Lemma 3] proved properties (65), (66) and (67) of Lemma 2. Thus what remains for us is to provide proof that the remaining event (68) has high probability.
In [22], large deviation results for dependent random variables were employed, but the underlying probability employed in codeword selection was the same as the one used by us, so that our findings connect seamlessly.

We become a bit more concrete now. Let $p \in \mathcal{P}_0^n(\mathcal{A})$, $q \in \mathcal{P}_0^n(\mathcal{S})$. Throughout, we will attempt to twist and tweak asymptotic quantities such that they are calculated with respect to the random variables (S, X, Z) defined via $\mathbb{P}((S, X, Z) = (s, x, z)) := p(x)q(s)v(z|x, s)$. Since the distribution of (S, X, Z) is so important, we label it by p_{SXZ}. The variable p will remain fixed, and q will always denote a type corresponding to one of the choices of James.
The proof will require us to draw codewords at random. As stated already, we adapt this procedure to the one chosen in [22]. This is done as follows: We define the random variables $X_{kl\gamma}$ $(1 \leqslant k \leqslant K,\ 1 \leqslant l \leqslant L,\ 1 \leqslant \gamma \leqslant \Gamma)$ by $\mathbb{P}(X_{kl\gamma} = x^n) := \frac{1}{|T_p|} \mathbb{1}_{T_p}(x^n)$ for all $k \in [K],\ l \in [L]\ \gamma \in [\Gamma]$ and $x^n \in \mathcal{X}^n$, where K, L, Γ are natural numbers. We write $\mathbf{x}_{kl\gamma}$ for the realizations of the variable $X_{kl\gamma}$, instead of $x^n_{kl\gamma}$. The random variable $\mathbf{X} := (\mathbf{X}_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ is distributed such that each $\mathbf{X}_{kl\gamma}$ is independent of $\mathbf{X}_{k'l'\gamma'}$ if $(k, l, \gamma) \neq (k', l', \gamma')$. The realizations of $(\mathbf{X}_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ are written $\mathbf{x}$. We use the projections $\pi_{kl\gamma}$ defined by $\pi_{kl\gamma}(\mathbf{x}) := \mathbf{x}_{kl\gamma}$. Further

projections as e.g. $\mathbf{x}_\gamma := \pi_\gamma(\mathbf{x}) := (\mathbf{x}_{kl\gamma})_{k,l=1}^{K,L}$ are defined wherever there is a need.
In order to enhance readability, we will not only omit the superscript n in our codewords, but from time to time we will also write statements like $\forall s^n$, property P holds. Then, it is understood that P holds for all $s^n \in \mathcal{S}^n$.
When calculating expectations of any of the $X_{kl\gamma}$ we need no reference to k, l, γ due to independence of our random variables. We therefore add another random variable, X^n, distributed as $\mathbb{P}(X^n = x^n) = \frac{1}{|T_p|}\mathbb{1}_{T_p}(x^n)$ as well.
A first and crucial step for all that is to come in the proofs of the technical Lemmas 1 and 2 is to fix some $\delta > 0$ and $p \in \mathcal{P}(\mathcal{X})$ and define, for all $s^n \in \mathcal{S}^n$ and $z^n \in \mathcal{Z}^n$, the functions $\Theta_{s^n,z^n} : \mathcal{X}^n \to [0, b]$ (where $b := 2^{-n(H(Z|X,S)-f_2(n,\delta))}$ for some function f_2, as we will see soon) by

$$M(s^n, z^n) := \{x^n \in T_p : D(\bar{N}(\cdot|s^n, x^n, z^n)\|p_{SXZ}) \leqslant \delta\} \tag{131}$$

$$\Theta_{s^n,z^n}(x^n) := v^{\otimes n}(z^n|s^n, x^n)\mathbb{1}_{M(s^n,z^n)}. \tag{132}$$

In order to enhance readability, the dependence of both M and Θ on δ is suppressed here and in the following. All our proofs rely on a common strategy, which only deviates in one point: The codes which ensure reliable transmission. For non-symmetrizable AVWCs we rely on the work [22] and use the codes which are defined therein. This will be sufficient to obtain all the results that we claimed for the uncorrelated coding secrecy capacity.
The coding theorem for secret common randomness assisted secrecy capacity needs an additional definition of codes. This definition is as follows:
For every $n \in \mathbb{N}$, set $\Xi_n := \mathcal{P}_0^n(\mathcal{S})$. For every x^n, define (not necessarily disjoint) "decoding" sets by

$$\hat{D}_{x^n} := \bigcup_{\xi \in \Xi_n} T_{W_\xi,\delta}(x^n) \tag{133}$$

and for a collection $\mathbf{x}_\gamma := (x_{kl\gamma})_{k,l=1}^{K,L}$ of codewords with fixed value of γ set

$$D(\mathbf{x}_\gamma)_{kl} := \hat{D}_{x_{kl\gamma}} \bigcap \left(\bigcup_{k' \neq k} \bigcup_{l' \neq l} \hat{D}_{x_{k'l'\gamma}} \right)^{\complement}. \tag{134}$$

This defines the code $\mathcal{K}_n$. This definition allows the decoder to decode the randomization index l as well, an approach which works for AVWCs and compound (wiretap) channels with convex state sets via the minimax theorem. Note that this code will only ensure reliable transmission if Γ is sufficiently large.
In order to deliver a joint treatment of the subject it makes sense to define the following events, where we implicitly assume a functional dependence $\delta = \delta(\tau)$ that will be specified more exactly later during our proofs. The sets $E_3, \ldots, E_5$ depend only on τ, whereas E_1 depends also on $\mathfrak{W}$

and E_2 on $\mathfrak{V}$.

$$E_1 := \left\{ \mathbf{x} | \forall s^n, z^n, k : \ \frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L} \Theta_{s^n,z^n}(\mathbf{x}_{kl\gamma}) \in \left[(1 \pm 2^{-n\cdot\tau/4}) \mathbb{E}\Theta_{s^n,z^n} \right] \right\} \tag{135}$$

$$E_2 := \left\{ \mathbf{x} | \min_{s^n} \frac{1}{\Gamma} \sum_{\gamma=1}^{\Gamma} d_{s^n}(\mathcal{K}_\gamma) \geqslant 1 - 2 \cdot 2^{-n\delta/4} \right\}. \tag{136}$$

$$E_3 := \left\{ \mathbf{x} | \max_{\gamma, x^n, s^n} |\{(k,l) : (x^n, \mathbf{x}_{kl\gamma}, s^n) \in T_{\bar{N}(\cdot|x^n,\mathbf{x}_{kl\gamma},s^n)}\}| \leqslant 2^{n(|R - I(\mathbf{x}_{kl\gamma};x^n,s^n)|^+ + \tau)} \right\} \tag{137}$$

$$E_4 := \left\{ \mathbf{x} | \max_{\gamma, s^n} |\{(k,l) : I(\mathbf{x}_{kl\gamma}; s^n) > \tau\}| \leqslant K \cdot L \cdot 2^{-n\cdot\tau} \right\} \tag{138}$$

$$E_5 := \left\{ \mathbf{x} | \max_{\gamma, s^n} \left| \left\{ (k,l,\gamma) : \begin{array}{l} \text{There is } (k',l',\gamma') \neq (k,l,\gamma) \text{ such that} \\ I(\mathbf{x}_{kl\gamma}; \mathbf{x}_{k'l'\gamma'}, s^n) - |R - I(\mathbf{x}_{kl\gamma}; s^n)|^+ > \tau \end{array} \right\} \right| \leqslant K \cdot L \cdot 2^{-n\cdot\tau/2} \right\} \tag{139}$$

The average success probability $d_{s^n}(\mathcal{K}_\gamma)$ was defined in Definition 3. The events E_3, E_4, E_5 are proven to have high probability in [22] (actually, their proof is valid for $|\Gamma| = 1$ but can be extended to arbitrary $|\Gamma|$ by simple union bounds, which leads to the following statement:

Lemma 5 (Cf. [22])**.** *There is $c' > 0$ such that, if $\mathfrak{W}$ is non-symmetrizable, we have that*

$$\mathbb{P}(E_3 \cap E_4 \cap E_5) \geqslant 1 - \Gamma \cdot \exp(-2^{n\cdot c'}) \tag{140}$$

The bound in Lemma 5 is trivial whenever $\Gamma > \exp(2^{n\cdot c'})$. In the applications intended here, the maximal scaling of Γ with n will be exponential, so that nontrivial bounds arise.
Our main effort in the following will be to show that a similar bound is true for $\mathbb{P}(E_1)$ and $\mathbb{P}(E_2)$ under the right conditions on K, L and Γ. With respect to these conditions, any of the intersections $E_i \cap \ldots \cap E_j$ will then have very high probability as well.
For the proofs of both Lemma 1 and 2 it will be of importance to control the amount of information which leaks out to Eve. This will require us to prove that a careful random choice of codewords will be provably secure, and this is the main content of the following Lemma (which contains statements concerning the message transmission capabilities of the common randomness assisted codes defined in (133) and (134) as well).

Lemma 6. *Let $K, L, \Gamma \in \mathbb{N}$. Let the random variable $\mathbf{X}$ be as described above. Then for every $\tau > 0$ and $\beta > 0$ there is a $\delta > 0$ and and $N \in \mathbb{N}$ such that for all $n \geqslant N$ and types $p \in \mathcal{P}_0^n(\mathcal{X})$, the following statements are true:*

1. *If $\frac{1}{n}\log(L \cdot \Gamma) \geqslant E(p) + \tau$ and $\min_{x:p(x)>0} p(x) \geqslant \beta$, then $\mathbb{P}(E_1) \geqslant 1 - 2 \cdot |\mathcal{S} \times \mathcal{X} \times \mathcal{Z}|^n \cdot \exp(-2^{n\cdot\tau/6})$.*
2. *If $\frac{1}{n}\log(K \cdot L) \leqslant B(p) - \delta - 2 \cdot f_1(\sqrt{2\cdot\delta})$ then $\mathbb{P}(E_2) \geqslant 1 - \exp(n \cdot \log(|S|) - \Gamma \cdot 2^{-n\delta})$.*
3. *For every $\beta > 0$, $|\mathcal{X}|$, $|\mathcal{S}|$ and $|\mathcal{Z}|$, a functional dependence between δ and τ can be chosen such that $\lim_{\tau\to 0} \delta(\tau) = 0$.*

The number N depends on $|\mathcal{X}|$, $|\mathcal{S}|$, $|\mathcal{Z}|$ as well as on p (via the quantity $\beta := \min_{x\in\mathcal{X}:p(x)>0} p(x)$) and on δ.

Proof. Some of the statements we wish to prove here are not about the full random variable $\mathbf{X} = (X_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ but only about exponentially many parts of it. We do therefore feel the need to write a few lines concerning our strategy of proof. We adopt the usual point of view that $\mathbf{X}$ somehow generates matrices of codewords. In the special case treated here it will be convenient to think of realizations of $\mathbf{X}$ as a list of Γ matrices, all of which describe a code-book and each of these code-books uses the index l solely for making Eve obfuscated, while k is used to transmit messages. The fact that γ is known to both the sender and the receiver lets the receiver adapt his decoder appropriately, while Eve only sees the average over all code-books. The effective randomness used for obfuscation of Eve is therefore $L \cdot \Gamma$.
Before making this more precise, we need additional notation:
As stated already, the projections $\pi_{kl\gamma} : (\mathcal{X}^n)^{KL\Gamma} \to \mathcal{X}^n$ project onto the copy of X^n corresponding to k, l, γ, such that $\pi_{kl\gamma}(\mathbf{X}) = X_{kl\gamma}$. Accordingly, π_k are the projections mapping $\mathbf{X}$ to $\mathbf{X}_k := (X_{kl\gamma})_{l,\gamma=1}^{L,\Gamma}$.
The trick will be to first understand how to embed statements concerning only certain projections of $\mathbf{X}$ into the whole random selection process. The idea is to proceed as follows:
Take any set of functions $g_1, \dots, g_M : \mathcal{X}^n \to [0, b']$. Then for all $k \in [K]$,

$$\mathbb{P}(\frac{1}{\Gamma \cdot L} \sum_{l,\gamma=1}^{L,\Gamma} g_m(\pi_{kl\gamma}(\mathbf{X})) \notin [(1 \pm \epsilon)\mathbb{E}g_m]) = \mathbb{P}(\frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L,\Gamma} g_m(\pi_{l\gamma}(\mathbf{X}_k)) \notin [(1 \pm \epsilon)\mathbb{E}g_m]), \quad (141)$$

where the left hand side is a probabilistic statement about $\mathbf{X} = (X_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ and the right hand side is a statement about the random variables $\mathbf{X}_k = (X_{kl\gamma})_{l,\gamma=1}^{L,\Gamma}$. Thus by the usual Chernoff bound Lemma 4 we have

$$\mathbb{P}\left(\exists\ m,k:\ \frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L,\Gamma} g_m(\pi_{l\gamma}(\mathbf{X}_k)) \notin [(1 \pm \epsilon)\mathbb{E}g_m]\right) \leqslant 2 \cdot M \cdot K \cdot \exp\left(-\frac{L \cdot \Gamma \cdot \epsilon^2 \cdot \min_m \mathbb{E}g_m}{3 \cdot b'}\right). \quad (142)$$

Another crucial connection in what is to follow is that for all z^n, x^n and s^n we have (using the abbreviation $N(\cdot) := N(\cdot|s^n, x^n, z^n)$ and $r(z|x,s) := N(s,x,z)/N(s,x|s^n,x^n)$):

$$\begin{aligned} v^{\otimes n}(z^n|s^n,x^n) &= 2^{n \cdot \sum_{s,x,z} \bar{N}(s,x,z) \log v(z|s,x)} && (143)\\ &= 2^{n \cdot (\sum_{s,x,z} \bar{N}(s,x,z)(\log \frac{v(z|s,x)p(x)q(s)}{\bar{N}(s,x,z)} + \log \frac{\bar{N}(s,x,z)}{p(x)q(s)}))} && (144)\\ &= 2^{n \cdot (-D(\bar{N}(\cdot|s^n,x^n,z^n)\|p_{XSZ}) + \sum_{s,x,z} \bar{N}(s,x,z) \log(\frac{\bar{N}(s,x|s^n,x^n) \cdot r(z|x,s)}{p(x) \cdot q(s)})} && (145)\\ &= 2^{n \cdot (-D(\bar{N}(\cdot|s^n,x^n,z^n)\|p_{XSZ}) + D(\bar{N}(\cdot|s^n,x^n)\|p \otimes q) - H(\hat{Z}|\hat{S},\hat{X}))}, && (146) \end{aligned}$$

where $\hat{S}\hat{X}\hat{Z}$ is distributed according to $\bar{N}$ (note that without loss of generality we may assume that $p, q > 0$ here and in the following lines, since otherwise we could simply erase a symbol from the alphabet $\mathcal{X}$ or $\mathcal{S}$).

Proof of property 1 of Lemma 6: Let $n \in \mathbb{N}$. Replace M with $\mathcal{S}^n \times \mathcal{Z}^n$ and the functions g_m with the Θ_{s^n,z^n}'s. We let $\delta > 0$ be arbitrary for the moment. Using equation (143)

and the fact that the relative entropy is never negative it can be seen that each Θ_{s^n,z^n} obeys

$$\Theta_{s^n,z^n}(x^n) = 2^{n\cdot(-D(\bar{N}(\cdot|s^n,x^n,z^n)\|p_{XSZ})+D(\bar{N}(\cdot|s^n,x^n)\|p\otimes q)-H(\hat{Z}|XS))}\mathbb{1}_{M(s^n,z^n)}(x^n) \tag{147}$$

$$\leqslant 2^{-n\cdot(H(\hat{Z}|\hat{S}\hat{X})-D(\bar{N}(\cdot|s^n,x^n)\|p\otimes q))}. \tag{148}$$

This bound does obviously still depend on x^n. But if $x^n \in M(s^n, z^n)$ then the distribution of $\hat{S}\hat{X}\hat{Z}$ has the following important feature: by Pinsker's inequality, we have

$$\|\bar{N} - p_{SXZ}\|_1 \leqslant \sqrt{2\delta}. \tag{149}$$

Setting $f_2(\delta) := 2 \cdot f_1(\sqrt{2\delta}) + \delta$, an application of Lemma 11 from the appendix together with monotonicity of the relative entropy then yields

$$\forall\ x^n \in \mathcal{X}^n : \qquad \Theta_{s^n,z^n}(x^n) \leqslant 2^{-n\cdot(H(Z|XS)-f_2(\delta))}. \tag{150}$$

Here, f_1 is defined setting $\mathcal{A} = \mathcal{S} \times \mathcal{X} \times \mathcal{Z}$. This justifies our choice of b. Note that the definition of Θ together with the monotonicity of $D(\cdot\|\cdot)$ ensures that the empirical distribution $\bar{N}(\cdot|x^n, s^n)$ is almost product ($\bar{N}(\cdot,\cdot|x^n, s^n) \approx p(\cdot) \cdot \bar{N}(\cdot|s^n)$) and that this property was vital in the derivation of the results contained in [22], whereas it may not be strictly necessary here (but does lead to a valid strategy of proof, nonetheless).
In order to apply the Chernoff bound we also need to calculate the expectation of each Θ_{s^n,z^n}, and for that matter it will be important to obtain a tight enough lower bound on $|M(s^n, z^n)|$: According to Lemma 9 from the appendix (set $\mathcal{A} = \mathcal{X}$ and $\mathcal{B} = \mathcal{S} \times \mathcal{Z}$ there) we have

$$|M(s^n, z^n)| \geqslant 2^{n(H(\hat{X}|\hat{S}\hat{Z})-f_C(n))}. \tag{151}$$

We are now almost ready to give a lower bound on the expectation of Θ_{s^n,z^n}. Be aware that s^n of type q and z^n remain fixed quantities for the moment. From monotonicity of the relative entropy and Pinsker's inequality applied together with Lemma 11 it follows that we can estimate

$$x^n \in M(s^n, z^n) \qquad \Rightarrow \qquad v^{\otimes n}(z^n|s^n, x^n) \geqslant 2^{-n(H(Z|X,S)+2\delta+f_1(\sqrt{2\delta}))}. \tag{152}$$

It then follows that, if $M(s^n, z^n) \neq \varnothing$, we have the estimate

$$\mathbb{E}\Theta_{s^n,z^n} = \frac{1}{|T_p|}\sum_{x^n\in M(s^n,z^n)} v^{\otimes n}(z^n|x^n, s^n) \tag{153}$$

$$\geqslant 2^{-n(H(Z|X,S)+2\delta+f_1(\sqrt{2\delta}))} \cdot 2^{n(H(X)-f_C(n))}|M(s^n, z^n)|. \tag{154}$$

Estimate (149) together with the continuity of entropy yields (see [20, Lemma 2.7])

$$M(s^n, z^n) \neq \varnothing \qquad \Rightarrow \qquad |M(s^n, z^n)| \geqslant 2^{n(H(X|S,Z)+f_C(n)+f_1(\sqrt{2\delta}))}. \tag{155}$$

We define $m : \mathcal{S}^n \times \mathcal{Z}^n \to \{0, 1\}$ by $m(s^n, z^n) = 1$ if $M(s^n, z^n) \neq \varnothing$ and $m(s^n, z^n) = 0$ else. It then follows that for all large enough $n \in \mathbb{N}$

$$\mathbb{E}\Theta_{s^n,z^n} \geqslant m(s^n, z^n) \cdot 2^{-n(H(Z|S)-f_3(n,\delta))}, \tag{156}$$

where $f_3(\delta) := 4(\delta + f_1(\sqrt{2\delta})$. For our random variable $\mathbf{X}$ this can be used as follows: via the Chernoff bound,

$$\mathbb{P}(\exists k,\ s^n,\ z^n :\ \frac{1}{L\cdot\Gamma}\sum_{l,\gamma=1}^{L,\Gamma}\Theta_{s^n,z^n}(\pi_{kl\gamma}(\mathbf{X}))\notin[(1\pm\epsilon)\mathbb{E}\Theta_{s^n,z^n}]) \tag{157}$$

$$\leqslant 2\cdot|\mathcal{S}\times\mathcal{X}\times\mathcal{Z}|^n\cdot\exp\left(-\epsilon^2\cdot L\cdot\Gamma\cdot\frac{\min_{s^n,z^n}\mathbb{E}\Theta_{s^n,z^n}}{3\cdot b}\right) \tag{158}$$

$$= c(n)\cdot\exp\left(-\epsilon^2\cdot\Gamma\cdot L\cdot\frac{\min_{s^n,z^n}\mathbb{E}\Theta_{s^n,z^n}}{3\cdot b}\right), \tag{159}$$

on account of the same argument that we used in equations (141) and (142) and with the obvious definition of $c(n)$. Now we have to plug in the asymptotic behaviour of $L\cdot\Gamma$, ϵ and b. If $m(s^n,z^n)=0$ then the statement is trivial. We set $f(\delta) := f_2(\delta)+f_3(\delta)$, $E(p) := \max_q I(p;V_q)$ and let $\frac{1}{n}\log L\cdot\Gamma \geqslant E(p)+\tau$ for some $\tau>0$. Note that, no matter what the distribution of S (which depends on the choice s^n of James!), we have $E(p)-I(X;Z|S)\geqslant 0$. Therefore,

$$\frac{\epsilon^2}{3}\cdot L\cdot\Gamma\cdot\frac{\mathbb{E}\Theta_{s^n,z^n}}{b} \geqslant m(s^n,z^n)\frac{\epsilon^2}{3}\cdot 2^{n(E(p)+\tau-H(Z|S)+H(Z|X,S)-f_2(\delta)-f_3(\delta))} \tag{160}$$

$$= m(s^n,z^n)\frac{\epsilon^2}{3}\cdot 2^{n(E(p)+\tau-I(Z;X|S)-f(\delta))} \tag{161}$$

$$\geqslant m(s^n,z^n)\frac{\epsilon^2}{3}\cdot 2^{n(E(p)+\tau-E(p)-f(\delta))} \tag{162}$$

$$= m(s^n,z^n)\frac{\epsilon^2}{3}\cdot 2^{n(\tau-f(\delta))}. \tag{163}$$

Upon choosing $\epsilon = 2^{-n\cdot\alpha}$ we get a doubly exponential decay of the probability in equation (157) if $0>\tau-2\alpha-f(\delta)$, and since $\lim_{\delta\to 0}f(\delta)=0$ there is a combination of $\delta>0$, $\tau>0$ such that for $\alpha=\tau/6$ and all large enough $n\in\mathbb{N}$ we have

$$\mathbb{P}\left(\exists\ k,s^n,z^n :\ \frac{1}{L\cdot\Gamma}\sum_{l,\gamma=1}^{L,\Gamma}\Theta_{s^n,z^n}(\mathbf{x}_{kl\gamma})\notin[(1\pm 2^{-n\tau/6})\mathbb{E}\Theta_{s^n,z^n}]\right)\leqslant c(n)\cdot\exp(-2^{n\cdot\tau/6}). \tag{164}$$

It is clear that this defines a dependence $\delta=\delta(\tau)$ and that $\lim_{\tau\to 0}\delta(\tau)=0$ and $\delta(\tau)>0$ for all (small enough) τ. A specific choice that we will use here is $\delta(\tau)=\tau$.

Proof of statement 2 of Lemma 6: We will need Ahlswede's robustification technique.

Lemma 7 ([3, 4])**.** *If a function $f:\mathcal{S}^n\to[0,1]$ satisfies*

$$\sum_{s^n\in\mathcal{S}^n} f(s^n)q(s_1)\cdots q(s_n)\geqslant 1-\varepsilon \tag{165}$$

for all $q\in\mathcal{P}_0^n(\mathcal{S})$ and some $\varepsilon\in[0,1]$, then

$$\frac{1}{n!}\sum_{\pi\in\Pi_n} f(\pi(s^n))\geqslant 1-3\cdot(n+1)^{|\mathcal{S}|}\cdot\varepsilon. \tag{166}$$

We will in the following make use of the codes $\mathcal{K}_\gamma$ which defined the set E_2.
We would like to use the Chernoff bound for the variable Γ, so we have to control the expectation for each fixed γ. Note that the construction of codes is such that it is independent from γ, so this will not turn into a hopeless case if we draw an independent number Γ of realizations of above codes. We go as follows: First associate to any given choice $\mathbf{x}_\gamma = (\mathbf{x}_{kl\gamma})_{k,l=1}^{K,L}$ of codewords the corresponding code $\mathcal{K}(\mathbf{x}_\gamma)$ as defined in equations (133) and (134). Then, for every s^n and $\gamma \in [\Gamma]$, define the success probability of that code via

$$d_{s^n}(\mathbf{x}_\gamma) := \sum_{k,l=1}^{K,L} \frac{1}{K \cdot L} w_{s^n}(D(\mathbf{x}_\gamma)_{kl}|\mathbf{x}_{kl\gamma}). \tag{167}$$

We then have for each fixed γ

$$\begin{aligned}
\mathbb{E} d_{s^n}(\mathbf{X}_\gamma) &= \mathbb{E} \frac{1}{K \cdot L} \sum_{k,l=1}^{K,L} w_{s^n}(D(\mathbf{X}_\gamma)_{kl}|\mathbf{X}_{kl\gamma}) && (168)\\
&\geqslant \mathbb{E} \frac{1}{K \cdot L} \sum_{k,l=1}^{K,L} \left(w_{s^n}(\hat{D}_{\mathbf{X}_{kl\gamma}}|\mathbf{X}_{kl\gamma}) - w_{s^n}(\bigcup_{k' \neq k} \bigcup_{l' \neq l} \hat{D}_{\mathbf{X}_{k'l'\gamma}}|\mathbf{X}_{kl\gamma}) \right) && (169)\\
&\geqslant \sum_{x^n \in T_p} \frac{1}{|T_p|} w_{s^n}(\hat{D}_{x^n}|x^n) - K \cdot L \cdot \sum_{x^n, \hat{x}^n \in T_p} \frac{1}{|T_p|^2} w_{s^n}(\hat{D}_{x^n}|\hat{x}^n). && (170)
\end{aligned}$$

Now observe that $\pi(T_p) = T_p$ for every $\pi \in S_n$ and that, for all $\pi \in S_n$, x^n, y^n and s^n we have $w_{s^n}(\pi(y^n)|\pi(x^n)) = w_{\pi^{-1}(s^n)}(y^n|x^n)$. In addition to that, $\hat{D}_{\pi(x^n)} = \pi(\hat{D}_{x^n})$, so that we can write

$$\mathbb{E} d_{s^n}(\mathbf{X}_\gamma) \geqslant \frac{1}{n!} \sum_{\pi \in S_n} w_{s^n}(\hat{D}_{\pi(x^n)}|\pi(x^n)) - K \cdot L \cdot \sum_{x^n, \hat{x}^n \in T_p} \frac{1}{|T_p|^2} w_{s^n}(\hat{D}_{x^n}|\hat{x}^n). \tag{171}$$

By Lemma 2.3 and equation (2.1) in [20], the density $\frac{1}{|T_p|}\mathbb{1}_{T_p}$ satisfies

$$\begin{aligned}
\frac{1}{|T_p|}\mathbb{1}_{T_p} &\leqslant (n+1)^{|\mathcal{X}|} 2^{-n \cdot H(p)} \mathbb{1}_{T_p} && (172)\\
&= (n+1)^{|\mathcal{X}|} p^{\otimes n} \mathbb{1}_{T_p} && (173)\\
&\leqslant (n+1)^{|\mathcal{X}|} p^{\otimes n}. && (174)
\end{aligned}$$

Setting $\mathrm{pl}(n) := (n+1)^{|\mathcal{X}|}$, we use this to further develop our bound as follows:

$$\begin{aligned}
\mathbb{E} d_{s^n}(\mathbf{X}_\gamma) &\geqslant \frac{1}{n!} \sum_{\pi \in S_n} w_{s^n}(\hat{D}_{\pi(x^n)}|\pi(x^n)) - K \cdot L \cdot \mathrm{pl}(n) \cdot \sum_{x^n \in T_p} \frac{1}{|T_p|} w_p^{\otimes n}(\hat{D}_{x^n}|s^n) && (175)\\
&= \frac{1}{n!} \sum_{\pi \in S_n} w_{s^n}(\hat{D}_{\pi(x^n)}|\pi(x^n)) - K \cdot L \cdot \mathrm{pl}(n) \cdot \sum_{\pi \in S_n} \frac{1}{n!} w_p^{\otimes n}(\hat{D}_{x^n}|\pi(s^n)), && (176)
\end{aligned}$$

where $x^n \in T_p$ is arbitrary and $w_p(y|s) = \sum_{x \in \mathcal{X}} p(x) w(y|s,x)$ according to our definition in equation (11). By carrying out the same estimate as in equation (172) for the distribution $\frac{1}{|T_q|}\mathbb{1}_{T_q}$ induced by the type q of s^n and setting $\mathrm{pl}_2(n) := (n+1)^{2 \cdot \max\{|\mathcal{X}|, |\mathcal{S}|\}}$ we get (note here

that $w_{p\otimes q}(y) := \sum_{s,x} q(s)\cdot p(x)\cdot w(y|s,x)$ defines, according to our convention, a probability distribution on $\mathcal{P}(\mathcal{Y})$ which is identical to $W(p\otimes q)$)

$$\mathbb{E}d_{s^n}(\mathbf{X}_\gamma) \geqslant \frac{1}{n!}\sum_{\pi\in S_n} w_{s^n}(\hat{D}_{\pi(x^n)}|\pi(x^n)) - K\cdot L\cdot \mathrm{pl}_2(n)\cdot w_{p\otimes q}^{\otimes n}(\hat{D}_{x^n}) \tag{177}$$

$$= \frac{1}{n!}\sum_{\pi\in S_n} w_{\pi(s^n)}(\hat{D}_{x^n}|x^n) - K\cdot L\cdot \mathrm{pl}_2(n)\cdot w_{p\otimes q}^{\otimes n}(\hat{D}_{x^n}) \tag{178}$$

$$\geqslant \frac{1}{n!}\sum_{\pi\in S_n} w_{\pi(s^n)}(\hat{D}_{x^n}|x^n) - K\cdot L\cdot \mathrm{pl}_2(n)\cdot \max_{\xi\in\Xi_n} w_{p\otimes q}^{\otimes n}(T_{W_\xi,\delta}(x^n)). \tag{179}$$

It is now the time to apply Ahlswede's robustification technique. For the fixed but arbitrary $x^n\in T_p$ define f by fixing all its values $f(s^n)$ via $f(s^n) := w_{s^n}(\hat{D}_{x^n}|x^n)$. Then by Lemma 7 we get

$$\mathbb{E}d_{s^n}(\mathbf{X}_\gamma) \geqslant 1-(n+1)^{|\mathcal{S}|}\max_{\xi\in\Xi_n} w_\xi^{\otimes n}(\hat{D}^{\complement}_{x^n}|x^n) - K\cdot L\cdot \mathrm{pl}_2(n)\cdot \max_{\xi\in\Xi_n} w_{p\otimes q}^{\otimes n}(T_{W_\xi,\delta}(x^n)) \tag{180}$$

$$\geqslant 1-\mathrm{pl}_2(n)\left(\max_{\xi\in\Xi_n} W_\xi^{\otimes n}(T_{W_\xi,\delta}(x^n)^{\complement}|x^n) + \right. \tag{181}$$

$$\left. + K\cdot L\cdot \max_{\xi\in\Xi_n} w_{p\otimes q}^{\otimes n}(T_{W_\xi,\delta}(x^n))\right) \tag{182}$$

$$\geqslant 1-\mathrm{pl}_2(n)\left(2^{-n\cdot\delta/2} + K\cdot L\cdot \max_{\xi\in\Xi_n}\prod_{x\in\mathcal{X}} w_{p\otimes q}^{\otimes n\cdot p(x)}(T_{W_\xi(x),\delta})\right). \tag{183}$$

The last term in above estimate deserves special attention. Following the lines of proof of Lemma 3 in [9] (which was originally proven in [42]) we see that

$$D(\bar{N}(\cdot|y^n)\|W_\xi(p)) = D(\sum_x p(x)\bar{N}_x(\cdot|y^n)\|W_\xi(p)) \tag{184}$$

$$\leqslant \sum_x p(x) D(N_x(\cdot|x^n,y^n)\|W_\xi(p)) \tag{185}$$

$$= D(\bar{N}(\cdot|x^n,y^n)\|W_\xi(p)\otimes p) \tag{186}$$

$$\leqslant \delta. \tag{187}$$

It follows that for each $\xi\in\Xi_n$ we have by Lemma 11 that

$$W_{p\otimes q}^{\otimes n}(T_{W_\xi,\delta}(x^n)) \leqslant |T_{W_\xi,\delta}(x^n)|\max_{y^n\in T_{W_\xi,\delta}(x^n)} w_{p\otimes q}^{\otimes n}(y^n) \tag{188}$$

$$\leqslant |T_{W_\xi,\delta}(x^n)|\max_{y^n\in T_{W_\xi,\delta}(x^n)} 2^{-n(D(\bar{N}(\cdot|y^n)\|W(p\otimes q))+H(\bar{N}(\cdot|y^n)))} \tag{189}$$

$$\leqslant |T_{W_\xi,\delta}(x^n)|2^{-n(H(W_\xi(p))-f_1(\sqrt{2\cdot\delta}))}. \tag{190}$$

We further estimate that for the distribution $p_{XY,\xi}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ defined via $p_{XY}(x,y) := p(x)w_\xi(y|x)$ we have

$$|T_{W_\xi,\delta}(x^n)| \leqslant \max_{y^n: D(\bar{N}(\cdot|x^n,y^n)\|p_{XY,\xi})\leqslant\delta} |\{\hat{y}^n : N(\cdot|\hat{y}^n,x^n) = N(\cdot|y^n,x^n)\}| \tag{191}$$

$$\leqslant \max_{y^n: D(\bar{N}(\cdot|x^n,y^n)\|p_{XY,\xi})\leqslant\delta} 2^{n\cdot H(\hat{Y}|\hat{X})} \tag{192}$$

$$\leqslant 2^{n\cdot\sum_x p(x)H(W_\xi(\delta_x))+f_1(\sqrt{2\cdot\delta})}, \tag{193}$$

by Lemma 9 and Lemma 11. We can now re-insert this estimate into our original problem and obtain

$$\mathbb{E}d_{s^n}(\mathbf{X}_\gamma) \geqslant 1 - \mathrm{pl}_2(n)\left(2^{-n\cdot\delta} + K\cdot L\cdot 2^{-n(\min_\xi I(p;W_\xi)-2f_1(\sqrt{2\cdot\delta}))}\right) \tag{194}$$

$$\geqslant 1 - \mathrm{pl}_2(n)\left(2^{-n\cdot\delta/2} + K\cdot L\cdot 2^{-n(\min_q I(p;W_q)-2\cdot f_1(\sqrt{2\cdot\delta}))}\right) \tag{195}$$

$$\geqslant 1 - \mathrm{pl}_2(n)\left(2^{-n\cdot\delta/2} + 2^{-n\cdot\delta/2}\right) \tag{196}$$

$$\geqslant 1 - 2^{-n\cdot\delta/4} \tag{197}$$

for all large enough $n\in\mathbb{N}$, since

$$K\cdot L \leqslant 2^{n(\min_q I(p;W_q)-\delta-2\cdot f_1(\sqrt{2\cdot\delta}))} \leqslant 2^{n(\min_\xi I(p;W_\xi)-\delta-2\cdot f_1(\sqrt{2\cdot\delta}))} \tag{198}$$

by assumption and since $\Xi_n \subset \mathcal{P}(\mathcal{S})$. Observe that this lower bound is entirely independent from the choice of $s^n\in\mathcal{S}^n$. It now follows from the Chernoff bound Lemma 4 that

$$\mathbb{P}(\forall s^n:\ \frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma} d_{s^n}(\mathcal{K}_\gamma) \leqslant (1-\epsilon)\mathbb{E}d_{s^n}(\mathcal{K})) \leqslant |\mathcal{S}|^n\cdot\exp(-\Gamma\cdot\epsilon^2\cdot\mathbb{E}d_{s^n}(\mathcal{K})/3) \tag{199}$$

$$\leqslant \exp(n\cdot\log(|S|) - \Gamma\cdot\epsilon^2\cdot(1-2^{-n\cdot\delta/4})/3). \tag{200}$$

Choose $\epsilon = 2^{-n\cdot\delta/4}$ to obtain the statement.

Proof of statement 3 in Lemma 6: The proof of this statement follows from the proof of statement 1 where the functional dependence $\tau\mapsto\delta(\tau)$ is specified. □

4.5 Proof of Lemma 1

Proof of Lemma 1. We know from Lemma 6 that (if $\frac{1}{n}\log(K\cdot L) \leqslant B(p) - \delta - 2\cdot f_1(\sqrt{2\cdot\delta})$ for some $\delta>0$ and n is large enough)

$$\mathbb{P}(E_2) \geqslant 1 - \exp(n\cdot\log(|\mathcal{S}|) - \Gamma\cdot\epsilon^2\cdot(\frac{1}{3} - 2^{-n\cdot\delta/2})). \tag{201}$$

Stepping away from the goal of proving Lemma 1 we see that there are two possible routes which diverge from here. One is to make Γ as small as possible, the other will be to exploit large numbers Γ. We will soon go on with the second approach and thereby prove Lemma 1, but first let us assume that we want Γ to be as small as possible (in an asymptotic sense of course). How can we achieve this? We take any sequence $(\epsilon_n)_{n\in\mathbb{N}}$ of numbers $\epsilon_n\in[0,1]$ which converges to zero. Depending on such a choice, we set $\Gamma_n = 3\cdot\log(|\mathcal{S}|^2)\frac{n}{\epsilon_n^2}(1-2^{-n\delta})$. It follows for the average success probability $d_{s^n}(\mathcal{K}_\gamma)$ as defined in Definition 3 that

$$\mathbb{P}(\ \forall s^n:\ \frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma} d_{s^n}(\mathcal{K}_\gamma) \leqslant (1-\epsilon)\mathbb{E}d_{s^n}(\mathcal{K})\ ) < 1, \tag{202}$$

proving the existence of a sequence of codes for which

$$\min_{s^n\in\mathcal{S}^n}\frac{1}{\Gamma_n\cdot K\cdot L}\sum_{\gamma,k,l=1}^{\Gamma_n,K,L} w_{s^n}(D_{k\gamma}|x_{k\gamma}) \geqslant 1-\epsilon_n \tag{203}$$

(whenever Γ_n scales asymptotically as $\Gamma_n \approx \frac{n}{\epsilon_n^2}$). If $\epsilon_n = n^{-\nu}$ for some small number $\nu > 0$ for example we get $\Gamma_n \approx \frac{n}{n^{-2\nu}} = n^{1+2\nu}$. This type of asymptotic scaling of common randomness has been observed several times now in the literature, and obviously raises the question whether $\Gamma_n = \text{const} \cdot n$ would be sufficient to guarantee asymptotically optimal performance, for some sufficiently large number const depending only on $|\mathcal{S}|$, for example.
We can now proceed our proof of Lemma 1 by using equation (200) together with Lemma 6 and a union bound: Let $\beta > 0$ and $\tau > 0$. From now on until the end of this proof, let $\delta = \delta(\tau)$. Let

$$\frac{1}{n}\log(\Gamma_n \cdot L_n) \geqslant E(p) + \tau, \qquad B(p) - \delta - 2 \cdot f_1(\sqrt{2 \cdot \delta}) \geqslant \frac{1}{n}\log(K_n \cdot L_n). \tag{204}$$

It then follows that for all large enough n it holds that

$$\mathbb{P}(E_1 \cap E_2) > 0. \tag{205}$$

Thus, there is a realization $\mathbf{x}$ of $\mathbf{X}$ such that for this particular realization we have

$$\forall s^n, z^n, k: \ \frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} \Theta_{s^n,z^n}(\mathbf{x}_{kl\gamma}) \in \left[(1 \pm 2^{-n\tau/4})\mathbb{E}\Theta_{s^n,z^n}\right] \tag{206}$$

$$\min_{s^n \in \mathcal{S}^n} \frac{1}{\Gamma}\sum_{\gamma=1}^{\Gamma} d_{s^n}(\mathcal{K}_\gamma) \geqslant 1 - 2 \cdot 2^{-n\delta/2} \tag{207}$$

Further, for every $k \in [K_n]$ we have (setting $\Delta(s^n, z^n, x^n) := \Theta_{s^n,z^n}(x^n)$ for all s^n, z^n and x^n)

$$\left\| \frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \mathbb{E}v_{s^n} \right\|_1 \tag{208}$$

$$\leqslant \left\| \frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} (v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \Delta(s^n, \cdot, \mathbf{x}_{kl\gamma}) \right\|_1 + \left\| \frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} \Delta(s^n, \cdot, \mathbf{x}_{kl\gamma}) - \mathbb{E}\Delta(s^n, \cdot, X^n) \right\|_1 \tag{209}$$

$$+ \left\| \mathbb{E}(v_{s^n}(\cdot|X^n) - \mathbb{E}\Delta(s^n, \cdot, X^n) \right\|_1 \tag{210}$$

$$\leqslant \frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} \left\| v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \Delta(s^n, \cdot, \mathbf{x}_{kl\gamma}) \right\|_1 + 2^{-n\cdot\tau/4} + \mathbb{E}\left\| v_{s^n}(\cdot|X^n) - \Delta(s^n, \cdot, X^n) \right\|_1 \tag{211}$$

where the first inequality is due to the triangle inequality of $\|\cdot\|_1$ and the second one due to the specific probabilistic choice of $\mathbf{x}$, especially the validity of (206). We now use the definition of Θ_{s^n,z^n} in order to derive bounds on the remaining quantities: for every $x^n \in T_p$ we have

$$\left\| v_{s^n}(\cdot|x^n) - \Delta(s^n, \cdot, x^n) \right\|_1 = \sum_{z^n : D(\bar{N}(\cdot|s^n,x^n,z^n)\|p_{SXZ}) > \delta} v^{\otimes n}(z^n|s^n, x^n) \tag{212}$$

$$= v^{\otimes n}(T_{V,\delta}(s^n, x^n)^{\complement}|s^n, x^n) \tag{213}$$

$$\leqslant 2^{-n\cdot\delta/2}, \tag{214}$$

for all large enough n. Thus

$$\frac{1}{L \cdot \Gamma}\sum_{l,\gamma=1}^{L,\Gamma} \left\| v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \Delta(s^n, \cdot, \mathbf{x}_{kl\gamma}) \right\|_1 + \mathbb{E}\left\| v_{s^n}(\cdot|X^n) - \Delta(s^n, \cdot, X^n) \right\|_1 \leqslant 2 \cdot 2^{-n\delta/2} \tag{215}$$

for all large enough $n \in \mathbb{N}$ so that we ultimately get (uniformly in $k \in [K]$) the bound

$$\frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L,\Gamma} \|v_{s^n}(\cdot|\mathbf{x}_{kl\gamma}) - \Delta(s^n, \cdot, \mathbf{x}_{kl\gamma})\|_1 \leqslant 2 \cdot 2^{-n\delta/2} + 2^{-n\tau/4} \leqslant 2^{-n\cdot\nu(\tau)}, \tag{216}$$

for all large enough n and setting $\nu(\tau) := \min\{\delta(\tau), \tau\}/5$ (note that $\nu(\tau) = \tau/5$ is a valid choice). □

4.6 Proof of Theorems 2, 3 and 4 (properties of C_{S})

Proof of Theorem 2. We give the proof of the properties of C_{S} in the same order as they were stated in the theorem:

1. This is clear from [27] where it was proven that symmetrizability makes it impossible to reach reliable transmission of messages.

2. The strategy of proof is to use Lemma 2 with $\Gamma = 1$. The reason for this is that, by assumption, $\mathfrak{W}$ is non-symmetrizable. Now, we know from Example 1 that this does not imply that every $\mathfrak{W} \circ U$ is non-symmetrizable as well. More precisely, to a given $r \in \mathbb{N}$ there may exist an alphabet $\mathcal{U}_n$, a $p \in \mathcal{P}(\mathcal{U}_n)$ and a channel $U_n \in C(\mathcal{U}, \mathcal{X}^n)$ such that

$$\min_{q \in \mathcal{P}(\mathcal{S}^r)} I(p; W_q \circ U_r) - \max_{s^r \in \mathcal{S}^r} I(p; V_{s^r} \circ U_r) \tag{217}$$

$$= \max_{p' \in \mathcal{P}(\mathcal{U}_r)} \max_{U'_r \in C(\mathcal{U}_r, \mathcal{X}^r} \min_{q \in \mathcal{P}(\mathcal{S}^r)} I(p'; W_q \circ U'_r) - \max_{s^r \in \mathcal{S}^r} I(p'; V_{s^r} \circ U'_r) \tag{218}$$

$$\geqslant C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) - \epsilon \tag{219}$$

but, additionally, $(W_{s^r} \circ \mathcal{U}_r)_{s^r \in \mathcal{S}^r}$ is symmetrizable. We provide here two approaches to deal with this problem: First, we will use the fact that $\mathfrak{W}$ is non-symmetrizable for transmission of a small number of messages that can be read by Eve but, since backwards communication from Eve to James is forbidden, are sufficient to counter any of the allowed jamming strategies.
Second, we will consider a variant of the optimization problem (4) where optimization of U'_r is restricted to maps of the form $U'_r = Id \otimes U''_{2,\dots,r}$ and we will prove that these restricted maps are asymptotically as good as those that are derived from the original problem when it comes to calculating capacity. However, these maps have the additional property that they cannot turn a non-symmetrizable AVC into a symmetrizable one.
Now let $r \in \mathbb{N}$ be arbitrary but fixed and p, U_r as above. Let $k, l \in \mathbb{N}$ be such that $n = k + l$ and $l = \lfloor \lambda \cdot n \rfloor$, where $\lambda \in (0, 1)$ is arbitrary but fixed for the moment. Then from [22, Lemma 5], if $\hat{K}$ satisfies the assumptions of Lemma 2 with L set to one based on the properties (65), (66) and (67) of the lemma.
So, on the grounds of 2 and of the results proven in [22], we see that for every $m' \in \mathbb{N}$, $r \in \mathbb{N}$ and $\delta > 0$, $p \in \mathcal{P}_0^{m'}(\mathcal{U}_r)$ (where $\mathcal{U}_r = [|\mathcal{X}|^r]$) and $U \in C(\mathcal{U}_r, \mathcal{X}^r)$ there exists a code $\mathcal{K} = (\mathcal{K}_m)_{m=1}^{\infty}$ such that for every $s^{r\cdot m} \in (\mathcal{S}^r)^m = \mathcal{S}^{r\cdot m}$ we have

$$\frac{1}{K'_k} \sum_{a=1}^{K'_k} \sum_{x^k} w_{s^k}(D'_a | \mathbf{x}'_a) \geqslant 1 - \epsilon_k, \tag{220}$$

where $\{\epsilon_k\}_{k\in\mathbb{N}} \subset [0,1]$, $\lim_{k\to\infty}\epsilon_k = 0$ and it may be assumed that $K'_k = l^3$. In addition to that we know from [38] that there exist codes for $(\mathfrak{W},\mathfrak{V})$ such that

$$\min_{s^l\in\mathcal{S}l}\frac{1}{\Gamma_l}\frac{1}{K''_l}\sum_{a,b=1}^{\Gamma_l,K''_l}\sum_{x^l\in\mathcal{X}^l} u_l(x^l|a,b)w_{s^l}(D''_{a,b}|\mathbf{x}''_{ab}) \geqslant 1-\delta_l, \tag{221}$$

where $\{\delta_l\}_{l\in\mathbb{N}} \subset [0,1]$, $\lim_{l\to\infty}\epsilon_l = 0$, $\Gamma_l = l^3$, $U_l \in C([\Gamma_l]\times[K''_l],\mathcal{X}^l)$ is stochastic pre-coding and $D_{a,b}\cap D_{a,b'} = \varnothing$ whenever $b\neq b'$ ($a\in[\Gamma_l]$ is used as common randomness in [38], whereas here we will substitute the messages that were sent on the first k channel uses for it. Note that the messages on the first k channel uses are not secure against Eve). In addition to that it holds

$$\lim_{l\to\infty}\frac{1}{l}\log K''_l = C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W},\mathfrak{V}) - \nu \tag{222}$$

for some arbitrarily small $\nu > 0$ and

$$\lim_{l\to\infty}\frac{1}{l}\max_{\gamma\in[\Gamma_l]}\max_{s^l\in\mathcal{S}^l} I(\mathfrak{K}''_l;\mathfrak{Z}_{s^l}|\mathfrak{b}_l = a) = 0. \tag{223}$$

The mutual information is evaluated on the random variables defined via

$$\mathbb{P}_{s^l}((\mathfrak{K}''_l,\mathfrak{Z}_{s^l},\mathfrak{b}_l) = (b,z^l,a)) := \frac{1}{\Gamma_l}\frac{1}{K''_l}\sum_{x^l\in\mathcal{X}^l} u_l(x^l|a,b)v(z^l|s^l,x^l). \tag{224}$$

We concatenate the two codes by defining new stochastic encodings $E_n \in C([K''_l],\mathcal{X}^n)$ via

$$e_n((x^k,x^l)|b) := \sum_{a=1}^{\Gamma_l}\delta_{\mathbf{x}_a}(x^k)u_l(x^l|a,b) \tag{225}$$

and new decoding sets via

$$D_b := \cup_a D'_a\times D''_{a,b}\subset\mathcal{X}^n. \tag{226}$$

It holds $D_b\cap D_{b'} = \cup_{a,a'}(D_a\times D_{a,b}\cap D_{a'}\times D_{a',b'}) = \varnothing$. We set $K_n := K''_l$, $\alpha_n := \epsilon_k$ and $\beta_n := \delta_l$ for the l satisfying $l = \lfloor\lambda\cdot n\rfloor$ and the k satisfying $k = n-l$. Then $\lim_{n\to\infty}\alpha_n = \lim_{n\to\infty}\beta_n = 0$. As a consequence of the Innerproduct Lemma in [2] we know that for every $s^n = (s^k,s^l)$ we have

$$\frac{1}{K_n}\sum_{b=1}^{K_n}\sum_{x^n\in\mathcal{X}^n} e_n(x^n|b)w(D_b|s^n,x^n) \geqslant \frac{1}{K_l}\sum_{a,b=1}^{\Gamma_k,K'_l}\sum_{x^l\in\mathcal{X}^n} u(x^l|a,b)w(D'_a|s^k,x^k)w(D''_{a,b}|s^l,x^l) \tag{227}$$

$$\geqslant 1-2\max\{\alpha_n,\beta_n\}. \tag{228}$$

That the messages $b\in[K_n]$ are also asymptotically secure in the sense that

$$\lim_{n\to\infty}\frac{1}{n}\max_{s^n\in\int^n} I(\mathfrak{K}_n;\mathfrak{Z}_{s^n}) \leqslant \lim_{l\to\infty}\frac{\lambda}{l}\max_{s^l\in\mathcal{S}^l} I(\mathfrak{K}''_l;\mathfrak{Z}_{s^l}|\mathfrak{b}_l) \tag{229}$$

$$= 0 \tag{230}$$

follows from independence of the distributions of the messages b and the values a of the common randomness as described in the inequalities from (48) to (57). Especially inequality (48) is valid since as a consequence of (223). The rate of the code is calculated as

$$\lim_{n\to\infty}\frac{1}{n}\log K_n = \lambda\left(C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W},\mathfrak{V})-\nu\right). \tag{231}$$

Since ν can be arbitrarily close to 0 and λ can be chosen arbitrarily close to 1 we have proven the desired result.

We now explain the second approach to proving statement 2. in Theorem 2. Here we aim to utilize the full power of Lemma 2 with $\Gamma = 1$. Our starting point are the distributions p and the channels U arising from the optimization (4) for fixed $r \in \mathbb{N}$. Note that, without loss of generality, $\mathcal{U}_r = \mathcal{X}^r$ for every $r \in \mathbb{N}$ in (4). Set, for every $r \in \mathbb{N}$,

$$C_r := \max_{p\in\mathcal{P}(\mathcal{X}^r)}\max_{U_r\in C(\mathcal{X}^r,\mathcal{X}^r)}\min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U_r) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U_r). \tag{232}$$

Let $r \in \mathbb{N}$ be arbitrary but fixed. For an arbitrary $\epsilon \geqslant 0$, let p and U_r be such that

$$C_r - \epsilon = \min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U_r) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U_r). \tag{233}$$

Now define $\tilde{U}_{r+1}$ by $\tilde{u}_{r+1}((x_1,\ldots,x_{r+1})|(x,u)) := \sum_{x'\in\mathcal{X}} u_r((x',x_2,\ldots,x_{r+1})|u)\delta_x(x_1)$ for all $x, x_1, \ldots, x_{r+1} \in \mathcal{X}$ and $u \in \mathcal{U}_r = \mathcal{X}^r$. Then it holds that

$$\begin{aligned} C_{r+1} &\geqslant \min_{q\in\mathcal{P}(\mathcal{S}^{r+1})} I(p\otimes\pi;W_q\circ U_{r+1}) - \max_{s^{r+1}\in\mathcal{S}^{r+1}} I(p\otimes\pi;V_{s^{r+1}}\circ U_{r+1}) && (234)\\ &\geqslant \min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p;W_q\circ U_r) - \max_{s^r\in\mathcal{S}^r} I(p;V_{s^r}\circ U_r) - \log|\mathcal{X}| && (235)\\ &= C_r - \epsilon - \log|\mathcal{X}|, && (236)\end{aligned}$$

where $\pi \in \mathcal{P}(\mathcal{X})$ is defined by $\pi(x) := |\mathcal{X}|^{-1}$ for all $x \in \mathcal{X}$. This latter estimate is due to the equality $I(p\otimes\pi;V_{s^{r+1}}\circ U_{r+1}) = I(p;V_{s^r}\circ U) + I(\pi;V_{s_{r+1}})$, the data processing inequality and the fact that for arbitrary channels $S \in C(\mathcal{A}\times\mathcal{B},\mathcal{C})$ and $T \in C(\mathcal{A}'\times\mathcal{B}',\mathcal{C}')$, as well as distributions $q \in \mathcal{S}(\mathcal{B}\times\mathcal{B}')$ with respective marginal distributions $q_B \in \mathcal{P}(\mathcal{B})$ and $q_{B'} \in \mathcal{P}(\mathcal{B}')$ and $p \in \mathcal{S}(\mathcal{A}\times\mathcal{A}')$ with respective marginal distributions $p_A \in \mathcal{P}(\mathcal{A})$ and $q_{A'} \in \mathcal{P}(\mathcal{A}')$ we have

$$\forall\ (a,b,c)\in\mathcal{A}\times\mathcal{B}\times\mathcal{C}: \qquad \sum_{a',c'}\sum_{b,b'} s(c|a,b)t(c'|a',b')p(a,a')q(b,b') = \sum_b q_B p_A t(c|a,b). \tag{237}$$

Since $\mathfrak{W}$ is non-symmetrizable we know that $\mathfrak{W}^{\otimes r}\circ\tilde{U}_r$ is non-symmetrizable for every $r \geqslant 2$. The reason for that is explained as follows: Let again S, T be channels as above. Assume that S is symmetrizable but T is not. Then $S\otimes T$ is non-symmetrizable. This can be seen by assuming the existence of a symmetrising map $Q \in C(\mathcal{A}\times\mathcal{A}',\mathcal{B}\times\mathcal{B}')$. The statement

$$\begin{aligned}&\forall(a_1,a_2,a_1',a_2')\in\mathcal{A}^2\times\mathcal{A}'^2:\\ &\qquad\sum_{b,b'} s(\cdot|a_1,b)t(\cdot|a_1',b')q(b,b'|a_2,a_2') = \sum_{b,b'} s(\cdot|a_2,b)t(\cdot|a_2',b')q(b,b'|a_1,a_1')\end{aligned} \tag{238}$$

would obviously imply for any fixed choice of (a_1, a_2) the statement

$$\forall(a_1',a_2')\in\mathcal{A}\times\mathcal{A}': \qquad \sum_{b'} t(\cdot|a_1',b')q_{B'}(b'|a_2,a_2') = \sum_{b'} t(\cdot|a_2',b')q_{B'}(b'|a_1,a_1'), \tag{239}$$

where $q_{B'}(b'|a_1,a_1') := \sum_b q(b,b'|a_1,a_1')$. This would be in contradiction to non-symmetrizability of T. Since $\tilde{U}_r = U_{r-1} \otimes Id$ we can thus conclude that $\mathfrak{W}^{\otimes r} \circ \tilde{U}_r$ is non-symmetrizable. We now proceed with the proof of Theorem 2.
With this approach we have evaded the problem that $\mathfrak{W}^{\otimes r} \circ U_r$ may well be symmetrizable (see our Example 1).
By [22, Lemma 4] non-symmetrizability of $\mathfrak{W}^{\otimes r} \circ \tilde{U}_r$ implies that it is possible to define a decoder according to [22, Definition 3], with $N = K \cdot L$ and $[N]$ replaced by $[K] \times [L]$. Since only the number of codewords and their type ever enters the proof it makes no difference whether we enumerate them by one index taken from $[N]$ or by two indices taken from $[K] \times [L]$. This decoder is proven to work reliably in [22, Lemma 5] (even with an exponentially fast decrease of average error), if $N = K \cdot L$ satisfies the assumptions of Lemma 2 based on the properties (65), (66) and (67) of the lemma.
So, on the grounds of Lemma 2 and of the results proven in [22], we see that for every $m \in \mathbb{N}$, $r \in \mathbb{N}\backslash\{1\}$ and $\delta > 0$, $p \in \mathcal{P}_0^m(\mathcal{X}^r)$ and $U \in C(\mathcal{X}^{r-1}, \mathcal{X}^{r-1})$ there exists a code $\mathcal{K} = (\mathcal{K}_m)_{m=1}^{\infty}$ such that for every $s^{r\cdot m} \in (\mathcal{S}^r)^m = \mathcal{S}^{r\cdot m}$ we have

$$\frac{1}{K_m}\frac{1}{L_m}\sum_{k,l=1}^{K_m,L_m}\sum_{x^{m\cdot r}} w_{s^{m\cdot r}}(D_{kl}|x^{m\cdot r})u^{\otimes m}(x^{m\cdot r}|u_{kl}) \geqslant 1-\epsilon_m, \tag{240}$$

where $\{\epsilon_m\}_{m\in\mathbb{N}} \subset [0,1]$, $\lim_{m\to\infty}\epsilon_m = 0$ and it holds that

$$\liminf_{m\to\infty}\frac{1}{m}\log(K_m \cdot L_m) \geqslant \min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p; W_q^m \circ \tilde{U}_r) - \delta \tag{241}$$

(the code we use here is defined by using the codewords $\mathbf{x}_{kl\gamma}$ together with the decoder from [22, Definition 3] defined for the AVC $\mathfrak{W}^{\otimes r} \circ \tilde{U}_r := (W_{s^r} \circ (U_{r-1} \otimes Id))_{s^r\in\mathcal{S}^r}$) and

$$\max_{s^r\in\mathcal{S}^r} I(p; V_{s^r} \circ \tilde{U}_r) + 2\delta \geqslant \liminf_{m\to\infty}\frac{1}{m}\log L_m \geqslant \max_{s^r\in\mathcal{S}^r} I(p; V_{s^r} \circ \tilde{U}_r) + \delta, \tag{242}$$

implying that for a sequence $(p_m)_{m\in\mathbb{N}}$ of choices for p_m converging to some p having a decomposition $p = p' \otimes \pi$ for $p' \in \mathcal{P}(\mathcal{X}^{r-1})$ being an optimal choice in the sense of (232) we get

$$\liminf_{m\to\infty}\frac{1}{m}\log K_m \geqslant \min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p; W_q \circ \tilde{U}_r) - \max_{s^r\in\mathcal{S}^r} I(p; V_{s^r} \circ \tilde{U}_r) - 3\delta \tag{243}$$

$$\geqslant C_{r-1} - \log|\mathcal{X}| - 3\delta. \tag{244}$$

Also, it is clear from the last part of Lemma 2 (equation (68)) together with [38, Lemma 20] that the codes employed here are asymptotically secure in the strong sense:

$$\limsup_{m\to\infty}\max_{s^{r\cdot m}} I(\mathfrak{K}_m; \mathfrak{Z}_{s^{r\cdot m}}) = 0. \tag{245}$$

We now wish to apply the code for the extended channel $(\mathfrak{W}^{\otimes r} \circ \tilde{U}_r, \mathfrak{V}^{\otimes r})$ to the original channel $(\mathfrak{W}, \mathfrak{V})$. Define values $t_n \in \{0, \ldots, r-1\}$ by requiring $n = m \cdot r + t_n$ for them to hold for some suitably chosen $m = m(n) \in \mathbb{N}$. This quantity satisfies $-1 + n/r \leqslant m(n) \leqslant n/r$. For every $n \in \mathbb{N}$ we then define new decoding sets by

$$\hat{D}_{kl} := D_{kl} \times \mathcal{Y}^{t_n} \tag{246}$$

and new randomized encodings by setting for some arbitrary but fixed x^{t_n}

$$E(\hat{x}^n|k) := \sum_{l=1}^{L} \frac{1}{L} u^{\otimes n}(x^{m\cdot r}|u_{kl}) \cdot \delta_{x^{t_n}}(\hat{x}^{t_n}). \tag{247}$$

From the choice of codewords and the decoding rule it is clear that this code is asymptotically reliable. The asymptotic number of codewords (mind that $\hat{K}_n = K_{m(n)}$) calculated and normalized with respect to n, is

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \hat{K}_n &= \liminf_{n\to\infty} \frac{1}{m(n)\cdot r + t_n} K_{m(n)} && (248)\\
&\geqslant \liminf_{n\to\infty} \frac{1}{r} \cdot \frac{1}{m(n)+1} K_{m(n)} && (249)\\
&= \liminf_{n\to\infty} \frac{1}{r} \cdot \frac{1}{m(n)} \cdot \frac{m(n)}{m(n)+1} \cdot K_{m(n)} && (250)\\
&= \frac{1}{r} \liminf_{n\to\infty} \frac{1}{m(n)} \cdot K_{m(n)} && (251)\\
&= \frac{1}{r} \left(C_{r-1} - 3\delta\right) && (252)\\
&= \frac{1}{r-1} \cdot \frac{r-1}{r} \left(C_{r-1} - \log|\mathcal{X}| - 3\delta\right). && (253)
\end{aligned}$$

In addition to that, the code is secure: For each $n \in \mathbb{N}$, the distribution of the input codewords and Eve's outputs is

$$\begin{aligned}
&\mathbb{P}(\mathfrak{K}_n = k, \mathfrak{Z}_{s^n} = z^n)\\
&\quad = \sum_{l=1}^{L} \frac{1}{L} \sum_{x^{r\cdot m}} \sum_{x^{t_n}} u^{\otimes m}(x^{r\cdot m}|u_{kl}) v^{\otimes r\cdot m}(z^{r\cdot m}|x^{r\cdot m}, s^{r\cdot m}) v^{\otimes t_n}(z^{t_n}|x^{t_n}, s^{t_n}) && (254)\\
&\quad = \mathbb{P}(\mathfrak{K}_n = k, \mathfrak{Z}_{s^{r\cdot m}} = z^{r\cdot m}) \cdot v^{\otimes t_n}(z^{t_n}|x^{t_n}, s^{t_n}). && (255)
\end{aligned}$$

This demonstrates that (uniformly in $s^n \in \mathcal{S}^n$ and since $\mathfrak{K}_n = \mathfrak{K}_m$ holds) we have

$$I(\mathfrak{K}_n; \mathfrak{Z}_{s^n}) = I(\mathfrak{K}_n; \mathfrak{Z}_{s^{r\cdot m}}) + 0 = I(\mathfrak{K}_m; \mathfrak{Z}_{s^{r\cdot m}}). \tag{256}$$

Since the right hand side of above equation goes to zero for n going to infinity and since $\lim_{r\to\infty} \frac{r-1}{r} = 1$ we see that the capacity C_{S} is lower bounded by $\lim_{r\to\infty} \frac{1}{r} C_r$. It is not an immediate consequence that this implies we can reach the capacity $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) = C^*(\mathfrak{W}, \mathfrak{V})$. Fortunately it has been proven in [38] that

$$C^*(\mathfrak{W}, \mathfrak{V}) = \lim_{r\to\infty} \frac{1}{r} \max_{p\in\mathcal{P}(\mathcal{U}_n)} \max_{U_n \in C(\mathcal{U}, \mathcal{X}^n)} \left(\min_{q\in\mathcal{P}(\mathcal{S}^r)} I(p; W_q \circ U) - \max_{s^r \in \mathcal{S}^r} I(p; V_{s^r} \circ U) \right) \tag{257}$$

holds. Thus $\lim_{r\to\infty} \frac{1}{r} C_r = C^*(\mathfrak{W}, \mathfrak{V})$. This finally implies the desired result. □

Proof of Theorem 3. If $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) = 0$, there is nothing to prove. Assume that $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) > 0$. It is evident that, in that case, $\mathfrak{W}$ is not symmetrizable. The function F defined in Definition 12 is continuous with respect to the Hausdorff distance (proving this statement is in complete

analogy as the corresponding part in the proof of Theorem 5 in [15]). Thus, if $F(\mathfrak{W}) > 0$, then there is an $\epsilon > 0$ such that for all $\mathfrak{W}'$ satisfying $d(\mathfrak{W}, \mathfrak{W}') < \epsilon$ we know that $F(\mathfrak{W}') > 0$ as well. Thus, every of these $\mathfrak{W}'$ is non-symmetrizable.
For some suitably chosen $\epsilon' < \epsilon$ we additionally know from Theorem 9 in [38] that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}', \mathfrak{V}) > 0$ for all those $\mathfrak{W}'$ for which $d(\mathfrak{W}, \mathfrak{W}') < \epsilon'$. But since Theorem 1 shows that $F(\mathfrak{W}') > 0 \Rightarrow C_{\mathrm{S}}(\mathfrak{W}', \mathfrak{V}) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}', \mathfrak{V})$ this implies that

$$C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) > 0 \; \forall \; \mathfrak{W}' : \; d(\mathfrak{W}, \mathfrak{W}') < \epsilon'. \tag{258}$$

Since from Theorem 3 we know that positivity of $C_{\mathrm{S}}(\mathfrak{W}', \mathfrak{V})$ ensures that it equals $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}', \mathfrak{V})$, and since the latter is continuous, we are done. □

Proof of Theorem 4. Again, we prove everything in the same order as it is listed in the theorem.

1. Let C_{S} be discontinuous in the point $(\mathfrak{W}, \mathfrak{V})$. By Theorem 3 we know that this can only be the case if $C_{\mathrm{S}}(\mathfrak{W}, \mathfrak{V}) = 0$. If in addition we have $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) = 0$ then we have, since $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ is continuous, that for every $\epsilon > 0$ there is $\delta > 0$ such that for all $(\mathfrak{W}_\delta, \mathfrak{V})$ satisfying $d(\mathfrak{W}_\delta, \mathfrak{W}) < \delta$ we have $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_\delta, \mathfrak{V}) \leqslant \epsilon$. Since since $C_{\mathrm{S,ran}}^{\mathrm{mean}} \geqslant C_{\mathrm{S}}$ this would imply that C_{S} is continuous as well, in contradiction to the assumption. Thus $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) > 0$. Of course this immediately implies that $\mathfrak{W}$ has to be symmetrizable, by property 2. This is, in turn, equivalent to $F(\mathfrak{W}) = 0$. The definition of F can be picked up from equation (58), its connection to symmetrizability is obvious from the definition. The notion of symmetrizability is explained in the introduction in equation (3). Clearly, if for all $\epsilon > 0$ and $\mathfrak{W}'$ satisfying $d(\mathfrak{W}, \mathfrak{W}') < \epsilon$ we would have $F(\mathfrak{W}') = 0$, then $C_{\mathrm{S}}(\mathfrak{W}', \mathfrak{V}')$ would be zero in a whole vicinity of $(\mathfrak{W}, \mathfrak{V})$. Thus for all $\epsilon > 0$ there has to be at least one $\mathfrak{W}_\epsilon$ such that $d(\mathfrak{W}, \mathfrak{W}_\epsilon) < \epsilon$ but $F(\mathfrak{W}_\epsilon) > 0$.
The reverse direction is basically established by using all our arguments backwards: For all $\epsilon > 0$, let there be at least one $\mathfrak{W}_\epsilon$ such that $d(\mathfrak{W}, \mathfrak{W}_\epsilon) < \epsilon$ but $F(\mathfrak{W}_\epsilon) > 0$. Let in addition to that $F(\mathfrak{W}) = 0$ but $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) > 0$. Since $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ is continuous, there is a $\delta > 0$ such that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}', \mathfrak{V}') > (1/2) \cdot C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \mathfrak{V}) =: \alpha$ whenever $d((\mathfrak{W}, \mathfrak{V}), (\mathfrak{W}', \mathfrak{V}')) < \delta$.
For every $\epsilon' \leqslant (1/2)\min\{\epsilon, \delta\}$ we can therefore deduce the following: It holds that $C_{\mathrm{S}}(\mathfrak{W}_{\epsilon'}, \mathfrak{V}) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_{eps'}, \mathfrak{V}) \geqslant \alpha > 0$ (since $F(\mathfrak{W}_{\epsilon'}) > 0$), but $C_{\mathrm{S}}(\mathfrak{W}_0, \mathfrak{V}) = 0$. Thus C_{S} is discontinuous in the point $(\mathfrak{W}, \mathfrak{V})$.

2. Let C_{S} be discontinuous in the point $(\mathfrak{W}, \mathfrak{V})$. By property 4 this implies that for all $\epsilon > 0$ there is $\mathfrak{W}_\epsilon$ such that $d(\mathfrak{W}, \mathfrak{W}_\epsilon) < \epsilon$ but $F(\mathfrak{W}_\epsilon) > 0$. If $\hat{\mathfrak{V}}$ is such that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}, \hat{\mathfrak{V}}) > 0$ then the pair $(\mathfrak{W}, \hat{\mathfrak{V}})$ fulfills all the points in the second of the two equivalent formulations in statement 4, and this implies that C_{S} is discontinuous in the point $(\mathfrak{W}, \hat{\mathfrak{V}})$. □

4.7 Proof of Lemma 2

Proof of Lemma 2. The proof is in many ways similar to the one for Lemma 1. As we know already that for some $c' > 0$ and all large enough $n \in \mathbb{N}$

$$\mathbb{P}(E_3 \cap E_4 \cap E_5) \geqslant 1 - \Gamma \cdot \exp(2^{-n \cdot c'}) \tag{259}$$

holds from [22], there is not much left to prove, as only $\mathbb{P}(E_1)$ needs to be controlled in order to get statement (68) of Lemma 2. We know from Lemma 6 that both

$$\mathbb{P}(E_1^{\complement}) \leqslant 2 \cdot |\mathcal{X} \times \mathcal{S} \times \mathcal{Z}|^n \cdot \exp(-2^{n \cdot \tau/2}), \tag{260}$$

if we choose $\delta = \delta(\tau)$. Keeping in mind that we already know from [22] that $\mathbb{P}(E_3 \cap E_4 \cap E_5) \geqslant 1 - \Gamma \cdot \exp(2^{n \cdot c'})$ we can combine all the previous to get the statement

$$\mathbb{P}(E_3 \cap \ldots \cap E_1) \geqslant 1 - (2 + \Gamma) \cdot \exp(-2^{n \cdot c''}), \tag{261}$$

for some $c'' > 0$ and for all large enough n. If Γ scales at most exponentially there will thus exist $N_0 \in \mathbb{N}$ such that for all $n \geqslant N_0$ there exists a choice $\mathbf{x} = (\mathbf{x}_{kl\gamma})_{k,l,\gamma=1}^{K,L,\Gamma}$ satisfying all conditions in Lemma 2 and, in addition, the estimate

$$\forall\ s^n, z^n, k: \ \frac{1}{L \cdot \Gamma} \sum_{l,\gamma=1}^{L,\Gamma} \Theta_{s^n,z^n}(\mathbf{x}_{kl\gamma}) \notin [(1 \pm 2^{-n\tau/4})\mathbb{E}\Theta_{s^n,z^n}]. \tag{262}$$

That this leads to secure transmission is proven exactly as in the proof of Lemma 1. The Lemma is thus proven. □

4.8 Proof of Theorem 5 (super-activation results)

We will divide this proof into three parts, each corresponding to its counterpart in Theorem 5.

Proof. **1.** Let us start with the "only if" statement. Clearly, if $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is symmetrizable then $C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = 0$. So, this part of the statement is proven.
If, on the other hand, $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is not symmetrizable and $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0$ then on account of Theorem 2, statement 1, we know that $C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0$.
This proves the first part of the Theorem.

2. In [16], Section VI, an explicit example of a pair $(\mathfrak{W}_i, \mathfrak{V}_i)_{i=1,2}$ has been given with the property that $\mathfrak{W}_1$ is symmetrizable, but $\mathfrak{W}_2$ is not. By elementary calculus, this implies that $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is non-symmetrizable.
Since this holds, our Theorem 2, statement 1, shows that the uncorrelated capacity of $(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2)$ equals its randomness-assisted capacity.
In [16] it was further shown that $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1, \mathfrak{V}_1) > 0$ and $C_{\mathrm{S}}(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$.

3. By assumption, $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$ but $C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{V}_1, \mathfrak{W}_2 \otimes \mathfrak{V}_2) > 0$. The former implies $C_{\mathrm{S}}(\mathfrak{W}_i, \mathfrak{V}_i) = 0$ $(i = 1, 2)$. If $\mathfrak{W}_1$ and $\mathfrak{W}_2$ were symmetrizable then clearly $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ would be symmetrizable and by [27] the message transmission capacity of $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ would be zero, implying $C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = 0$. If on the other hand either $\mathfrak{W}_1$ or $\mathfrak{W}_2$ are not symmetrizable then $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is not symmetrizable and this implies

$$C_{\mathrm{S}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_1 \otimes \mathfrak{W}_2, \mathfrak{V}_1 \otimes \mathfrak{V}_2) > 0, \tag{263}$$

where the equality is due to Theorem 2, part 1, and the lower bound is true by assumption.

4. We do again rely on Theorem 2. Let both $\mathfrak{W}_1$ and $\mathfrak{W}_2$ be symmetrizable. Then $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ is symmetrizable. Since by assumption $C_{\mathrm{S,ran}}^{\mathrm{mean}}$ shows no super-activation on the pair $(\mathfrak{W}_i, \mathfrak{V}_i)$ $(i = 1, 2)$ it follows that C_{S} cannot show super-activation as well. Thus at least one of the two AVCs has to be non-symmetrizable. Let without loss of generality this channel be $\mathfrak{W}_1$. If in addition $\mathfrak{W}_2$ would be non-symmetrizable, then $C_{\mathrm{S}}(\mathfrak{W}_i, \mathfrak{V}_i) = C_{\mathrm{S,ran}}^{\mathrm{mean}}(\mathfrak{W}_i, \mathfrak{V}_i)$ would hold for $i = 1, 2$ and since $\mathfrak{W}_1 \otimes \mathfrak{W}_2$ would be symmetrizable as well, we would additionally have

$C_{\mathrm{S}}(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2) = C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_1\otimes\mathfrak{W}_2,\mathfrak{V}_1\otimes\mathfrak{V}_2)$. But since $C^{\mathrm{mean}}_{\mathrm{S,ran}}$ shows no super-activation on the pair $(\mathfrak{W}_i,\mathfrak{V}_i)$ $(i=1,2)$ this cannot be. Thus again without loss of generality we have $\mathfrak{W}_2$ is symmetrizable.
Since we are talking about super-activation of C_{S}, it has to be that $C_{\mathrm{S}}(\mathfrak{W}_i,\mathfrak{V}_i)=0$ holds for $i=1,2$. But since $\mathfrak{W}_1$ is non-symmetrizable this requires that $C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_1,\mathfrak{V}_1)=0$ holds. If in addition we would have $C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_2,\mathfrak{V}_2)=0$ would hold than C_{S} could not be super-activated since $C^{\mathrm{mean}}_{\mathrm{S,ran}}$ cannot be super-activated by assumption. Thus $C^{\mathrm{mean}}_{\mathrm{S,ran}}(\mathfrak{W}_2,\mathfrak{V}_2)>0$. □

4.9 Proof of Lemma 3

We now prove Lemma 3: First and without loss of generality, we have $\mathcal{A}\subset\mathcal{A}'$. Let $\mathfrak{U}$ be symmetrizable. Let $Q\in C(\mathcal{A},\mathcal{R})$ be the symmetrizing channel, meaning that for all $a,a'\in\mathcal{A}$ the equality

$$(U\circ(Id\otimes Q))(a,a') = (U\circ(Id\otimes Q))(a',a) \tag{264}$$

holds true. It follows that for all $a,a'\in\mathcal{A}'$ it holds that

$$\begin{aligned}(U\circ(T\otimes QT))(a,a') &= \sum_{a'',a'''\in\mathcal{A}}\sum_{r\in\mathcal{R}} u(\cdot|a'',r)t(a''|a)q(r|a''')t(a'''|a') &(265)\\ &= \sum_{a'',a'''\in\mathcal{A}}\sum_{r\in\mathcal{R}} u(\cdot|a''',r)t(a''|a)q(r|a'')t(a'''|a') &(266)\\ &= (U\circ(T\otimes QT))(a',a). &(267)\end{aligned}$$

Thus, $\mathfrak{U}'$ is symmetrizable.

5 Appendix (auxiliary results and proofs)

Lemma 8 (Cf. [11])**.** *Let $p\in\mathcal{P}(\mathcal{X})$. For every $n\geqslant|\mathcal{X}|^2$, there is $p'\in\mathcal{P}_0^n(\mathcal{X})$ such that*

$$\|p-p'\|_1\leqslant\frac{2|\mathcal{X}|}{n} \tag{268}$$

and $p(x)=0$ implies $p'(x)=0$ for all $x\in\mathcal{X}$.

Proof of Lemma 8. Let $n\in\mathbb{N}$ be arbitrary. Set $\mathcal{X}':=\{x\in\mathcal{X}:p(x)>0\}$. From the next lines it will follow that, without loss of generality, we may assume $\mathcal{X}=\mathcal{X}'$. For sake of simplicity, assume again without loss of generality that $\mathcal{X}=\{1,\ldots,|\mathcal{X}|\}$ and that $p(|\mathcal{X}|)\geqslant 1/|\mathcal{X}|$. Choose $p'(i)$, for $i=1,\ldots,|\mathcal{X}|-1$, such that $|p'(i)-p(i)|\leqslant\frac{1}{n}$. Clearly, this is possible. Then necessarily

$p'(|\mathcal{X}|) = 1 - \sum_{i=1}^{|\mathcal{X}|-1} p'(i)$ and

$$\|p - p'\|_1 \leqslant \sum_{i=1}^{|\mathcal{X}|-1} \frac{1}{n} + |p'(|\mathcal{X}|) - p(|\mathcal{X}|)| \tag{269}$$

$$= \frac{|\mathcal{X}| - 1}{n} + |\sum_{i=1}^{|\mathcal{X}|-1} p(i) - p'(i)| \tag{270}$$

$$\leqslant \frac{|\mathcal{X}| - 1}{n} + \sum_{i=1}^{|\mathcal{X}|-1} |p(i) - p'(i)| \tag{271}$$

$$\leqslant \frac{2|\mathcal{X}|}{n}. \tag{272}$$

Of course, while all the $p'(i) \geqslant 0$ by construction if $i < |\mathcal{X}|$, this does not hold for $p'(|\mathcal{X}|)$. This is where we need the additional condition that $n \geqslant |\mathcal{X}|^2$:

$$p'(|\mathcal{X}|) = 1 - \sum_{i=1}^{|\mathcal{X}|-1} p'(i) \tag{273}$$

$$\geqslant 1 - \sum_{i=1}^{|\mathcal{X}|-1} p(i) - \frac{|\mathcal{X}| - 1}{n} \tag{274}$$

$$\geqslant p(|\mathcal{X}|) - \frac{|\mathcal{X}|}{n} \tag{275}$$

$$\geqslant \frac{1}{|\mathcal{X}|} - \frac{|\mathcal{X}|}{n} \tag{276}$$

$$\geqslant 0. \tag{277}$$

□

Lemma 9 (C.f. [19])**.** *Let $\hat{a}^n \in \mathcal{A}^n$ and $\hat{b}^n \in \mathcal{B}^n$. There exists a function $f_C : \mathbb{N} \to \mathbb{R}_+$ such that with $\hat{A}\hat{B}$ being distributed as $\mathbb{P}((\hat{A}, \hat{B}) = (a, b)) = \frac{1}{n} N(a, b|\hat{a}^n, \hat{b}^n)$ we have*

$$|\{a^n : N(\cdot|\hat{a}^n, \hat{b}^n) = N(\cdot|a^n, \hat{b}^n)\}| = 2^{n \cdot (H(\hat{A}|\hat{B}) - f_C(n))}. \tag{278}$$

The function f_C satisfies $\lim_{n\to\infty} f_C(n) = 0$.

The following Lemma is basically taken from [20]. It would generally be completely sufficient for proving all our statements in sufficient generality.

Lemma 10. *Let $D(p\|q) \leqslant \delta$. For the function $f_1 : [0, 1/2] \to \mathbb{R}_+$ defined by $f_1(x) := -\sqrt{x/2} \log(x|\mathcal{Z}|^2)$ we have that*

$$|H(p) - H(q)| \leqslant f_4(\delta). \tag{279}$$

Clearly, $\lim_{\delta\to 0} f_4(\delta) = 0$.

Note that $p(x) = 0$ implies $p'(x|s) = 0$ for all $s \in \mathcal{S}$, by construction.

Proof. From Pinsker's inequality we have $\|p-q\|_1 \leqslant \sqrt{2\delta}$ and, accordingly, by Lemma 2.7 in [20], $|H(p)-H(q)| \leqslant -\sqrt{2\delta}\log(\sqrt{2\delta}/|\mathcal{Z}|)$. □

We did however feel that it would be interesting to use a slightly more general version of Lemma 10, which led us to prove the following Lemma:

Lemma 11 (Continuity of conditional entropy with respect to averaged norm)**.** *Let $p \in \mathcal{P}(\mathcal{X})$ and channels $w, r : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{Z})$ be given such that*

$$\sum_{x\in\mathcal{X}} p(x)\|w(\cdot|x) - r(\cdot|x)\|_1 \leqslant \delta \leqslant 1. \tag{280}$$

Then

$$|H(w|p) - H(r|p)| \leqslant f_1(\delta), \tag{281}$$

where $f_1(\delta) := |\mathcal{Z}| \cdot h(\frac{\delta}{|\mathcal{Z}|})$.

Proof of Lemma 11. As in [20], set $\nu(t) := -t\log t$ and observe that ν is concave and satisfies $\nu(0) = \nu(1) = 0$. This brings with it the property that for all $s, \lambda \in [0,1]$ we have

$$\nu(\lambda \cdot a) \geqslant \lambda \cdot \nu(a), \qquad \nu(\lambda \cdot a + 1 - \lambda) \geqslant \lambda \cdot \nu(a). \tag{282}$$

We wish to obtain a meaningful bound on $|\nu(s) - \nu(t)|$ in terms of $|s-t|$. To this end, assume without loss of generality that $s \leqslant t$. Observe that this implies that $|t-s| = t-s$, so that both

$$\begin{aligned} \nu(|t-s|) + \nu(s) &= \nu(t \cdot \frac{t-s}{t}) + \nu(t \cdot \frac{s}{t}) && (283)\\ &\geqslant \frac{t-s}{t} \cdot \nu(t) + \frac{s}{t} \cdot \nu(t) && (284)\\ &= \nu(t) && (285) \end{aligned}$$

and with $\lambda := \frac{t-s}{1-s}$ satisfying $0 \leqslant \lambda \leqslant 1$ we have

$$\begin{aligned} \nu(1-|t-s|) + \nu(t) &= \nu(\lambda \cdot s + 1 - \lambda) + \nu(\lambda + (1-\lambda)\cdot s) && (286)\\ &\geqslant \lambda\nu(s) + (1-\lambda)\nu(s) && (287)\\ &= \nu(s), && (288) \end{aligned}$$

so that in total we get for every two number $s, t \in [0,1]$:

$$\begin{aligned} |\nu(t) - \nu(s)| &\leqslant \max\{\nu(|t-s|), \nu(1-|t-s|)\} && (289)\\ &\leqslant \nu(|t-s|) + \nu(1-|t-s|) && (290)\\ &= h(|t-s|) && (291) \end{aligned}$$

where h denotes the binary entropy. Then for every $(\epsilon_x)_{x\in\mathcal{X}} \in [-1,1]^{|\mathcal{X}|}$ and $(t_x)_{x\in\mathcal{X}} \in [0,1]^{|\mathcal{X}|}$ such that $t_x + \epsilon_x \in [0,1]$ for all $x \in \mathcal{X}$ we get:

$$\begin{aligned} |\sum_{x\in\mathcal{X}} p(x)(\nu(t_x) - \nu(t_x+\epsilon_x))| &\leqslant \sum_{x\in\mathcal{X}} p(x)|\nu(t_x) - \nu(t_x+\epsilon_x)| && (292)\\ &\leqslant \sum_{x\in\mathcal{X}} p(x)h(|\epsilon_x|) && (293)\\ &\leqslant h(\sum_{x\in\mathcal{X}} p(x)|\epsilon_x|). && (294) \end{aligned}$$

Then, we write $t_{xz} := w(z|x)$ and $\epsilon_{xz} := -w(z|x)+r(z|x)$. This leads to the bound we ultimately need:

$$|\sum_{x\in\mathcal{X}} p(x)H(w(\cdot|x)) - H(r(\cdot|x))| = |\sum_{z\in\mathcal{Z}}\sum_{x\in\mathcal{X}} p(x)(\nu(w(z|x)) - \nu(r(z|x)))| \tag{295}$$

$$\leqslant \sum_{z\in\mathcal{Z}} |\sum_{x\in\mathcal{X}} p(x)(\nu(t_{xz} - \nu(t_{xz}+\epsilon_{xz})| \tag{296}$$

$$\leqslant \sum_{z\in\mathcal{Z}} h(\sum_{x\in\mathcal{X}} p(x)|\epsilon_{xz}|) \tag{297}$$

$$\leqslant |\mathcal{Z}|\cdot h(\frac{1}{|\mathcal{Z}|}\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} p(x)|\epsilon_{xz}|) \tag{298}$$

$$= |\mathcal{Z}|\cdot h(\frac{1}{|\mathcal{Z}|}\delta) \tag{299}$$

□

Acknowledgements. J.N. wants to thank Prakash Narayan, Aylin Yener, Ebrahim MolavianJazi and Mohamed Nafea for fruitful and lively discussions. The authors are grateful to their unknown referees for helping them to increase the quality of the manuscript. This work was supported by the DFG via grant BO 1734/20-1 (H.B.) and by the BMBF via the grants 01BQ1050 and 16KIS0118 (H.B., J.N.).
Further funding (J.N.) was provided by the ERC Advanced Grant IRQUAT, the Spanish MINECO Project No. FIS2013-40627-P and the Generalitat de Catalunya CIRIT Project No. 2014 SGR 966.

References

[1] R. Ahlswede, "A Note on the Existence of the Weak Capacity for Channels with Arbitrarily Varying Channel Probability Functions and Its Relation to Shannon's Zero Error Capacity" *Ann. Math. Stat.*, Vol. 41, No. 3. (1970)

[2] R. Ahlswede, "Elimination of Correlation in Random Codes for Arbitrarily Varying Channels", *Z. Wahrscheinlichkeitstheor. Verw. Geb.* Vol. 44, 159-175 (1978)

[3] R. Ahlswede, "Coloring Hypergraphs: A New Approach to Multi-user Source Coding-II", *J. Comb. Inf. Syst. Sci.* Vol. 5, No. 3, 220-268 (1980)

[4] R. Ahlswede, "Arbitrarily Varying Channels with States Sequence Known to the Sender", *IEEE Trans. Inf. Th.* Vol. 32, 621-629, (1986)

[5] R. Ahlswede, N. Cai, "Correlated sources help the transmission over AVC", *IEEE Trans. Inf. Th.*, Vol. 43, No. 4, 1254-1255 (1997)

[6] R. Ahlswede, A. Winter, "Strong converse for identification via quantum channels", *IEEE Trans. Inf. Theory*, Vol. 48, No. 3, 569–579 (2002)

[7] S. Beigi, "A New Quantum Data Processing Inequality", *J. Math. Phys.*, Vol. 54, 082202 (2013)

[8] C.H. Bennett, P.W. Shor, J.A. Smolin, A.V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem" *IEEE Trans. Inf. Theory*, Vol. 48, No. 10, 2637-2655 (2002)

[9] I. Bjelaković, H. Boche, J. Somerfeld, "Secrecy Results for Compound Wiretap Channels", *Probl. Inf. Trans.*, Vol. 49, No. 1, 73-98 (2013)

[10] I. Bjelaković, H. Boche, J. Somerfeld, "Capacity Results for Arbitrarily Varying Wiretap Channels", *LNCS* Vol. 7777, 123-144 (2013)

[11] D. Blackwell, L. Breiman, A.J. Thomasian, "The Capacity of a Class of Channels" *Ann. Math. Stat.* Vol. 30, No. 4, 1229-1241 (1959)

[12] D. Blackwell, L. Breiman, A.J. Thomasian, "The capacities of certain channel classes under random coding", *Ann. Math. Stat.* Vol. 31, 558-567 (1960)

[13] M. Bloch, J.N. Laneman, "On the secrecy capacity of arbitrary wiretap channel," *Forty-Sixth Annual Allerton Conference, Allerton House, Illinois, USA* (2008)

[14] H. Boche, J. Nötzel, "Arbitrarily small amounts of correlation for arbitrarily varying quantum channels", *J. Math. Phys.* Vol. 54, 112202 (2013)

[15] H. Boche, J. Nötzel, "Positivity, discontinuity, finite resources, and nonzero error for arbitrarily varying quantum channels", *J. Math. Phys.*, Vol. 55, 122201 (2014)

[16] H. Boche, R.F. Schaefer, "Capacity Results and Super-Activation for Wiretap Channels With Active Wiretappers", *IEEE Trans. Inf. Forensic Secur.*, Vol. 8, No. 9, 1482-1496 (2013)

[17] H. Boche, R. F. Schaefer, "Arbitrarily Varying Wiretap Channels with Finite Coordination Resources", *Communications Workshops (ICC), 2014 IEEE International Conference on*, 746–751 (2014).

[18] H. Boche, R.F. Schaefer, H. Vincent Poor, "On the Continuity of the Secrecy Capacity of Compound and Arbitrarily Varying Wiretap Channels", *arXiv:1409.4752* (2014)

[19] I. Csiszar, "The Method of Types", *IEEE Trans. Inf. Theory* Vol. 44, No. 6, 2505-2523 (1998)

[20] I. Csiszar, J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, Cambridge, second edition (2011)

[21] I. Csiszár, J. Körner, "Broadcast Channels with Confidential Messages", *IEEE Trans. Inf. Theory*, Vol. IT-24, No. 3 (1978)

[22] I. Csiszár, P. Narayan, "The Capacity of the Arbitrarily Varying Channel Revisited: Positivity, Constraints", *IEEE Trans. Inf. Theory* Vol. 34, No. 2, 181-193 (1988)

[23] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel", *IEEE Trans. Inf. Theory*, Vol. 51, No. 1, 44-55 (2005)

[24] I. Devetak, P. Shor, "The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information", *Comm. Math. Phys.* Vol. 256, No. 2, 287-303 (2005)

[25] D.D. Dubhashi, A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*, Cambridge University Press (2012)

[26] A. El Gamal, Y.H. Kim, *Network Information Theory* Cambridge University Press (2012)

[27] T. Ericson, "Exponential Error Bounds for Random Codes in the Arbitrarily Varying Channel", *IEEE Trans. Inf. Th.*, Vol. 31, No. 1, 42-48 (1985)

[28] P. Gacs, J. Koerner, "Common information is far less than mutual information", *Probl. Control Inf. Th.*, Vol. 2, No. 2, 149-162, (1973)

[29] X. He, A. Khisti, A. Yener, "Mimo multiple access channel with an arbitrarily varying eavesdropper: Secrecy degrees of freedom", *IEEE Trans. Inf. Theory*, Vol. 59, No. 8, 4733-4745 (2013)

[30] W. Kang, N. Liu, "Wiretap Channel with Shared Key", *IEEE Inf. Theory Workshop - ITW 2010 Dublin* (2010)

[31] W. Kang, S. Ulukus, "A New Data Processing Inequality and Its Applications in Distributed Source and Channel Coding", *IEEE Trans. Inf. Theory*, Vol. 57, No. 1 (2011)

[32] J. Kiefer, J. Wolfowitz, "Channels with arbitrarily varying channel probability functions", *Information and Control* Vol. 5, 44-54 (1962)

[33] A. V. Kuznetsov, B. S. Tsybakov, "Coding in a memory with defective cells", *Probl. Inf. Transm.*, Vol. 10, No. 2, 52–60 (1974)

[34] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP Journal on Wireless Communications and Networking* (2008)

[35] A. Orlitsky, J.R. Roche, "Coding for Computing", *IEEE Trans. Inf. Theory* Vol. 47, No. 3, 903-917 (2001)

[36] G. Smith, J. Yard, "Quantum Communication With Zero-Capacity Channels", *Science* Vol. 321, 1812-1815 (2008)

[37] M. Wiese, "Multiple Access Channels with Cooperating Encoders" *Dissertation, München : Universitätsbibliothek der TU München* (2013)

[38] M. Wiese, J. Nötzel, H. Boche, "The Arbitrarily Varying Wiretap Channel—deterministic and correlated random coding capacities under the strong secrecy criterion", preprint, arXiv:1410.8078 (2014)

[39] H. S. Witsenhausen, "On sequences of pairs of dependent random variables", *SIAM J. Appl. Math.* Vol. 28, No. 1, (1975)

[40] A.D. Wyner, "The wire-tap channel," *The Bell System Tech. J.*, Vol. 54, No. 8, 1355–1387, (1975)

[41] A.D. Wyner "The Common Information of Two Dependent Random Variables", *IEEE Trans. Inf. Th.* Vol. IT-24, No. 2, 163-79 (1975)

[42] R.F. Wyrembelski, I. Bjelaković, T. Oechtering, H. Boche, "Optimal Coding Strategies for Bidirectional Broadcast Channels under Channel Uncertainty", *IEEE Trans. Commun.*, Vol. 58, No. 10, 2984–2994 (2010)